



TITLE:

保存された連続反応を用いた代謝
パスウェイのモジュール構造の同
定(Dissertation_全文)

AUTHOR(S):

武藤, 愛

CITATION:

武藤, 愛. 保存された連続反応を用いた代謝パスウェイのモジュール構造の同定. 京都大学, 2013, 博士(理学)

ISSUE DATE:

2013-09-24

URL:

<https://doi.org/10.14989/doctor.k17853>

RIGHT:

許諾条件により要旨は2013-10-01に公開

保存された連続反応を用いた
代謝パスウェイのモジュール構造の同定

武藤 愛

2013年

要旨

生物の物質代謝システムは、代謝反応によって結ばれた巨大な化合物ネットワークである。生物の物質生産能や分解能など生化学的特性のシステムの理解を助けるものとして、代謝ネットワークは最も重要な生体ネットワークの一つとされている。代謝ネットワークは酵素遺伝子によって世代間で受け継がれる遺伝的ネットワークであると同時に、代謝化合物を代謝反応によって繋ぐ化学的ネットワークでもある。化学的ネットワークの進化には、遺伝的な制限だけでなく、有機化学反応論に基づいた化学反応性という制限が存在すると考えられる。

代謝ネットワーク中には、異なる化合物の代謝パスウェイに共通した連続反応のパターンが共通して出現する例が知られている。このような連続反応は酵素の保存性によって見つけることができ、実際に酵素遺伝子のオペロン様構造との関連や、機能との結びつきが示されたものもある。一方で、保存された連続反応は代謝ネットワーク経路に存在する有機化学的禁制律を反映しているとも考えられる。本研究では、この連続反応パターンを化学反応モジュール（一つの単位として機能する連続反応）として代謝ネットワークから抽出し、機能との結びつきをアノテーション（注釈）として付与した代謝単位ライブラリーのリソースを作成することを目指した。

連続反応の保存性は、個々の研究者の知識や研究対象に頼ったボトムアップ的手法個別解析によって報告されてきた。本研究で私は、反応類似性スコアを定

義することで、酵素のアミノ酸配列に依存することなく、保存された連続反応を代謝ネットワークから抽出する手法を開発した。この手法により、急速に拡大する生化学ネットワークの知識から、トップダウン的アプローチによって、化学反応モジュールを網羅的に抽出することが可能になった。開発した代謝反応比較手法を用いて代謝ネットワークゲノムとパスウェイ中からの連続反応パターンの抽出を行った。代謝パスウェイの知識データベース KEGG PATHWAY に蓄積された代謝パスウェイネットワーク中の連続反応から長さ 2~8 の保存された連続反応を網羅的に抽出した。その結果、5,805 の代謝反応からなる代謝ネットワークから 3,016 の連続反応パターンが検出された。これらについて詳しく解析を行ったところ、化学反応モジュールとゲノム配列上のクラスター遺伝子との間に多くの対応関係が見られた。よって、これらの代謝の機能単位は、遺伝的にも一つの単位となって受け渡されている可能性が高い。また、アミノ酸など多くの代謝の基礎化合物の合成経路は、化学反応モジュールの組み合わせによって表せることがわかった。このことは、限られた数の反応列の組み合わせによってこれらの化合物代謝産物のバリエーションが生成されている可能性を示唆する。

得られた化学反応モジュールの情報は、これまで酵素の進化のみで語られてきた代謝ネットワーク経路の進化研究に、化学反応の進化という新しい観点を与えるものである。代謝ネットワーク経路がどのような有機化学的禁制律に沿って進化してきたかを探ることは、現在の生物の代謝ネットワーク経路予測に有益な指標を与えるだけでなく、生命の起源の探索に大きな手掛かりとなると考

えられる。本研究で得られた反応モジュールは、KEGG データベース (<http://www.genome.jp/kegg/>)の一部として公開されている。

目次

要旨.....	2
第 1 章 序論.....	9
1.1 生物と代謝系.....	9
1.2 代謝酵素.....	10
1.3 代謝パスウェイ.....	11
1.4 代謝パスウェイの進化.....	12
1.5 代謝パスウェイデータベース.....	15
1.6 代謝ネットワークのモジュール.....	19
1.7 パスウェイアライメント.....	22
1.8 本研究の目的.....	25
第 2 章 データセット.....	27
2.1 代謝パスウェイデータベース.....	27
2.2 反応ペア.....	28
2.3 KEGG atom type.....	30
2.4 RDM 化学構造変換表記法.....	32
2.5 反応クラス.....	35
第 3 章 結果 1 : 代謝反応クラス類似性比較手法の開発.....	36
3.1 フィンガープリント表記.....	36
3.2 RC エントリの類似性スコア.....	37
3.3 RCLASS エントリの類似性によるグループ化.....	41
3.4 RCLASS 保存配列パターンの抽出.....	41
第 4 章 結果 2 : 化学反応モジュール.....	44
4.1 抽出された保存反応列パターン.....	44
4.2 反応モジュールの出現頻度.....	44
4.3 反応モジュールの一般的特性.....	47
4.4 各化学反応モジュールの生化学的意味付け.....	47
4.4.1 2-オキシカルボン酸鎖の伸長.....	47

4.4.2	2-オキソカルボン酸の修飾.....	53
4.4.3	脂肪酸合成と β 酸化.....	59
4.4.4	微生物の生分解パスウェイにおける芳香族環分解.....	63
4.5	酵素遺伝子クラスターにコードされる反応モジュール.....	68
第 5 章 考察.....		72
5.1	逆反応の利用.....	72
5.2	モジュール間の包含関係.....	75
5.3	リン酸化と配糖化.....	76
5.4	一次代謝と二次代謝のモジュール情報の相違.....	79
5.5	EC 分類体系との比較.....	82
5.6	モジュール情報の利用.....	85
5.7	代謝ネットワークの進化.....	85
第 6 章 総括.....		89
謝辞.....		91
参考文献.....		92

図目次

図 1-1 代謝パスウェイの進化仮説.....	14
図 1-2 KEGG(Kyoto Encyclopedia of Genes and Genomes) データベース.....	17
図 1-3 KEGG PATHWAY map.....	18
図 1-4 ネットワークモジュール.....	20
図 1-5 EC 階層分類の問題点の例.....	24
図 2-1 反応ペアグラフ.....	29
図 2-2 RDM.....	33
図 3-1 RDM 表記のフィンガープリント表記への変換.....	37
図 3-2 パスウェイからの連続反応の抽出.....	42
図 4-1 代謝カテゴリ特異的に出現する反応モジュールの割合.....	45
図 4-2 2-オキソカルボン酸伸長モジュール.....	50
図 4-3 EC 1.1.1.42 と EC 1.1.1.41.....	51
図 4-4 アミノ酸生成モジュール.....	54
図 4-5 グルコシノレート生合成パスウェイ.....	55
図 4-6 カルボキシル基をアミノ基へ変換するモジュール.....	58
図 4-7 脂肪酸生合成モジュール.....	61
図 4-8 β 酸化モジュール.....	62
図 4-9 微生物分解パスウェイを構成する芳香族環分解モジュール.....	64
図 4-10 トルエン分解における好気性モジュール RM003 と 嫌気性モジュール RM015 の比較.....	66
図 4-11 反応モジュール RM001 に対応する <i>Pyrococcus furiosus</i> の二つの遺伝子クラスター.....	71
図 5-1 アミノ基-カルボキシル基変換モジュール.....	74
図 5-2 糖ヌクレオチド生合成モジュール.....	78
図 5-3 フェニルプロパノイド生合成.....	81
図 5-4 モノオキシゲナーゼの EC 分類と RCLASS の比較.....	84

表目次

表 2-1 KEGG atom type.....	31
表 3-1 フィンガープリントキー.....	36
表 3-2 KEGG atom type とフィンガープリントキーの対応.....	40
表 4-1 KEGG 代謝パスウェイ中に見つかった RCLASS 保存配列パターンの数.....	45
表 4-2 反応モジュールの一覧.....	46
表 4-3 2-オキソカルボン酸関連反応モジュール群.....	52
表 4-4 KEGG モジュールに対応する生分解反応モジュール.....	67
表 4-5 酵素遺伝子クラスターと対応する反応モジュール.....	70
表 5-1 最大の RC エントリを包含する EC 番号.....	83

第 1 章 序論

1.1 生物と代謝系

代謝は生命の最も基本的な要素の一つである。ある化合物を別の化合物へと変換する代謝能を有することは、自己複製能や外界と隔離された膜を持つことと並び、生命の定義の一つとされている。原始地球には単純な化合物しか存在しなかったと考えられており、現在地球上に存在する数十万種類もの化合物は、生物の代謝機能の発達と共に生み出されて来たものと思われる。我々が必須栄養素とするアミノ酸や糖、脂質なども、全て何らかの生物の代謝機能によって作られたものである。

代謝系とは、化学反応を通して生きた細胞中の全ての必須化合物を作り上げ、維持する化学システムである。ヒトの代謝系では 2 万ほど、植物では 20 万から 100 万もの代謝化合物が生成されており[1]、生体は多種多様な化合物が混ざり合い反応しあう巨大な系であるとも言える。しかしながら、代謝系と混合物との大きな違いは、系の中の恒常性が維持されていることにある。代謝系の化学反応はほぼ全てが代謝酵素により触媒されるが、代謝系の化学反応は酵素の量や活性を調節することによって、厳密に管理されている。代謝系は調節システムを持つ反応系であると言える。

代謝酵素はゲノム上にコードされ、世代を越えて伝えられることから、代謝系は化学反応の集合体であると同時に、遺伝的システムでもある。全ての生物種は祖先から綿々と受け継がれてきた酵素遺伝子のレパートリによって、異なる

生合成能や生分解能、環境への適応能を持つ。生物間の生息地の棲み分けや食性の違い、従属栄養に代表される依存関係のような、我々が目にすることのできる生物の様々な特性は、それぞれの生物種の代謝系の相違によりもたらされている。化学システムとしての代謝系と、遺伝システムとしての代謝系は、代謝系を考える上で欠かすことのできない二つの側面である。

1.2 代謝酵素

代謝系の化学反応のほとんどを担うのが、ゲノム上にコードされたタンパク質から成る代謝酵素である。代謝酵素はタンパク質表面に位置するアミノ酸残基によって基質化合物に特異的に結合し、反応を触媒する。酵素の階層分類を1961年に行っている The International Union of Biochemistry and Molecular Biology (IUBMB)の酵素委員会 (Enzyme Committee) によれば、酵素はおおまかに、オキシドリダクターゼ (酸化還元酵素)、トランスフェラーゼ (転移酵素)、ヒドラーゼ (加水分解酵素)、リアーゼ (付加脱離酵素)、イソメラーゼ (異性化酵素)、リガーゼ (合成酵素) の6つのECクラスに分けられる[2]。それぞれのクラスに属する酵素はさらに詳細に分類され、EC番号 (酵素番号) と呼ばれる数字列で酵素分類が表現される。代謝系に存在する酵素はこれらの反応のうちいずれかの触媒活性を持つが、これらの分類は必ずしも相互に排他的でないため、全ての酵素が1つの分類群にのみ所属するとは限らない。

分子系統学的観点からは、酵素はアミノ酸配列および構造の類似性によって、スーパーファミリーと呼ばれる同一の祖先遺伝子から生まれたとみなされる酵

素遺伝子群に分類される。タンパク質の構造分類データベース CATH データベース(CATH v3.5)[3]及び SCOP データベース(SCOP 1.75)[4]では現在 2,626 及び 1,962 のスーパーファミリーが定義されている。では、これらの分子系統的分類と、反応による酵素分類は一致するのであろうか。2003 年の Schmidt らによるスーパーファミリーと EC 番号の一致を調べた研究[5]によれば、ファミリー中の酵素の EC 番号が保存されていたスーパーファミリーは全スーパーファミリーの 36%にすぎず、多くのスーパーファミリーは複数の EC クラスに分類される酵素を含んでいた。また、全スーパーファミリーの 20%には全く異なる反応を触媒する酵素が存在した。つまり、同一の祖先遺伝子から生まれた酵素であっても、その触媒する化学反応や基質特異性が一致するとは限らない。

1.3 代謝パスウェイ

細胞中で起こる連鎖的な化学反応は代謝パスウェイと呼ばれる。代謝パスウェイでは、1つの化学反応の生成物が続く反応の基質となり、出発点の化合物は経路を通りながら段階的に修飾を受け、別の物質に変化する。一連の化学反応の途中で生じる様々な中間体は代謝中間体と呼ばれる。ほとんどの代謝反応は可逆的であるが、多くの代謝パスウェイは一方向に進むと考えられている。代謝パスウェイ中の化学反応は、多くの場合それぞれ個別の酵素が担っている。つまり、代謝パスウェイは出発点の化合物から最終産物を生み出すために、複数の酵素が協調してはたらく生産ラインであると言える。

バクテリアでは、代謝パスウェイ上で連続した反応を担う酵素群が、ゲノム上でも連続して位置している例が多く知られている。このようにゲノム上で連続して並ぶ遺伝子群は遺伝子クラスターと呼ばれる。遺伝子クラスターのうち、一つの転写因子によって同時に発現が制御される遺伝子群をオペロンと呼ぶ。オペロンや遺伝子クラスターを形成する遺伝子群は、代謝系において機能的関連性を持つ可能性が高い。

1.4 代謝パスウェイの進化

代謝パスウェイは基質が連鎖的な触媒反応を経て目的の化合物へと変化する一連の反応列であり、各反応を酵素が触媒する。このような経路は、基質を目的の産物へと変換するまでのひと揃いの代謝酵素群が揃って初めて機能する。それぞれの酵素が単独で存在しても、基質は目的化合物に辿りつけず、その細胞の生命活動に寄与しないため、酵素が単独で進化してきたとは考えにくい。代謝パスウェイがどのように進化してきたのかという問いについては、長年に渡り議論が続いている。

代謝パスウェイの進化仮説としては、これまで様々なモデルが提唱されている(図 1-1)。1945 年に Horowitz が提唱した **Retrograde model**[6]は、代謝パスウェイは後ろから前に向けて伸長してきたという説である。原始地球には様々な化合物が環境中に潤沢に存在し、基質にあたる化合物だけでなく基質に類似した化合物も豊富に存在したとする。目的の基質が全て反応に使われ、環境中から基質が得られなくなった後には、基質を使い切った酵素から、環境下にある別の

化合物を目的の基質に変換する酵素が **tandem duplication** により生まれ、上流の反応を行う酵素へと特化していく。それを繰り返すことで代謝パスウェイが拡大してきたとする。この仮説は、環境中には現在中間産物にあたる化合物が豊富に存在したという仮定に基づいており、原始地球における代謝パスウェイの初期進化のモデルとしては妥当とされるものの、現在の地球環境には当てはまらない。また、新規酵素が **tandem duplication** によって獲得されたならば、隣接する酵素は互いにホモログ遺伝子であるはずである。オペロンの発見により Horowitz (1965) は仮説を改め、任意のパスウェイの酵素をコードする遺伝子のクラスター構造は、代謝系初期において **tandem duplication** が起こったと考えることで説明されうるとした[7]。しかしながら Lawrence と Roth が 1996 年に発表したデータ解析の結果はこの可能性を否定している[8]。原核生物のゲノム上でのホモログ遺伝子のペアの隣接関係が保存されている例はわずかししか知られていない。[9]

代謝パスウェイデータベースの解析結果から最も支持されているのが、Jensen が 1976 年に提唱した **Patchwork model**[10]である。**Patchwork model** によれば、現在の代謝パスウェイは、幅広い基質に対し反応を行う酵素が特定の代謝パスウェイに特化してきた結果だとされる。このような基質特異性の低い酵素は低効率であるため、細胞の生命活動に通常はあまり寄与しない。しかし、基質の枯渇などによって他の代謝パスウェイが機能しなくなった場合に、その代替経路として機能するよう淘汰が働き、より高い特異性と効率的な反応を達成するようになる可能性は十分に考えられる。近年の全ゲノム比較研究から、代

謝酵素の遺伝子には遺伝子重複により生じたと考えられるものがかなりの割合で存在することが明らかになっている。このことは、共通の祖先遺伝子に由来する遺伝子群が別の代謝パスウェイにリクルートされることで代謝パスウェイが拡大してきたという **Patchwork model** に強い支持を与えている。但し、このモデルは既に利用可能な酵素群が存在したということを前提としているため、原始地球における代謝パスウェイの初期の進化には当てはまらない。

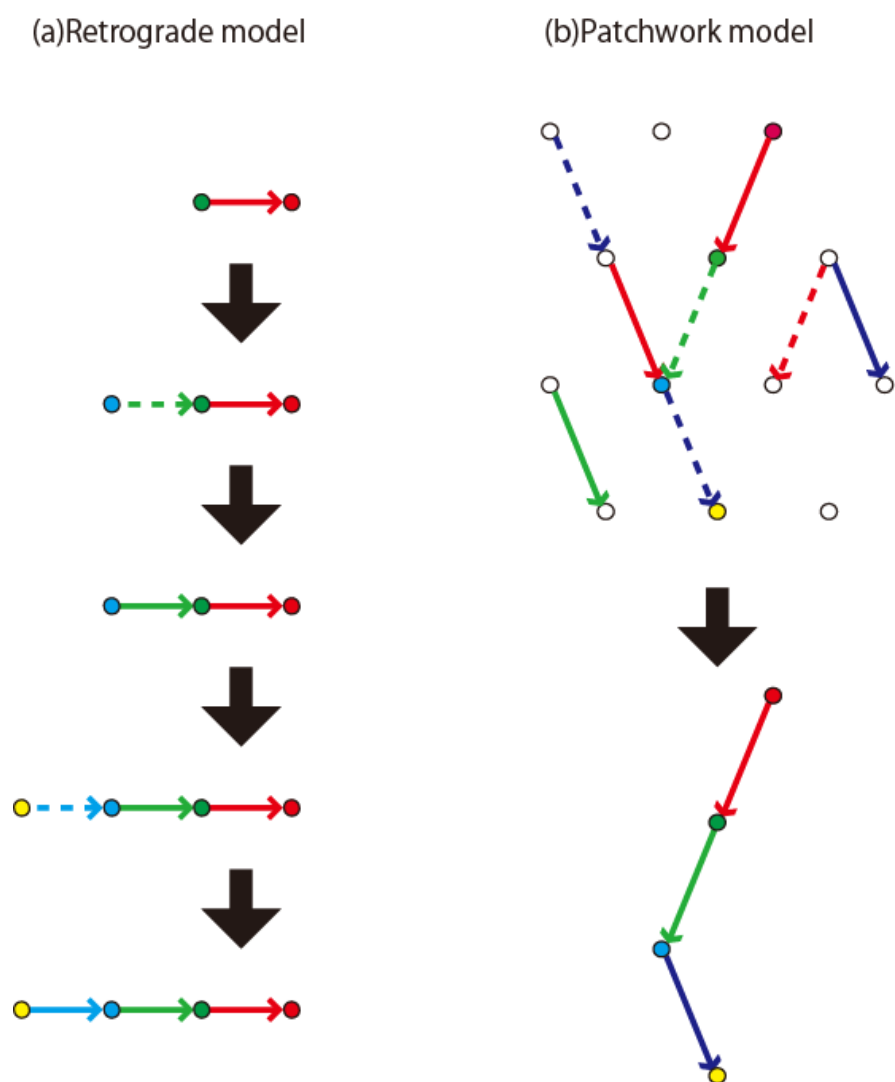



図 1-1 代謝パスウェイの進化仮説

Retrograde model や Patchwork model、及びその他の代謝ネットワークの進化モデルにおける代謝の進化は、これまで酵素の進化として語られてきた。代謝ネットワークの進化研究の主なアプローチは、現在の代謝系を担う酵素群の類縁関係から、どのような共通祖先から代謝酵素が分化してきたかを推測することであり、その結果どのような代謝反応が可能になったかということはネットワーク進化の「結果」として語られがちである。しかしながら、表現型の進化により直結しているのは、獲得した新規反応の方である。原始地球において嫌気性生物から進化して生まれた好気性生物は、酸化反応の利用により多くの有用な新規反応を獲得した[11]。このとき生まれた酵素反応は、酵素の分化という要因よりはむしろ、酸化反応というそれまで利用できなかった反応を、様々な基質に対して利用できるようになったことによって多様化したと言える。このように、酵素の進化と反応の進化は代謝系の進化の両輪であり、代謝ネットワークの進化を考える際には、酵素のネットワークだけではなく化学的ネットワークの進化についても考える必要があると考える。

1.5 代謝パスウェイデータベース

KEGG(Kyoto Encyclopedia of Genes and Genomes) データベース (図 1-2) は、遺伝子、ゲノム、化合物の情報を蓄積するだけでなく、オーソログ遺伝子（異なる生物種が持つ最終共通祖先が持っていた同一の遺伝子に由来する遺伝子）の情報や酵素遺伝子（および酵素タンパク質）とその反応との対応関係をリン

クし、横断的な解析を可能にするリソースを提供している[12]。それらのリソースの中でも最も利用されているのが、様々な生体内経路の情報を提供する **KEGG PATHWAY** データベース中の代謝パスウェイ情報である。**KEGG PATHWAY** では代謝化合物をノード（頂点）、代謝反応をエッジ（辺）として表す(図 1-3)。ユーザーはエッジに付与された酵素の情報をもとに、目的の生物種の代謝パスウェイを再構築して可視化することができる。代謝の再構築は通常、ゲノム中から既知の酵素のオーソログ遺伝子を見つけ出し、それらを代謝パスウェイのネットワークにマッピングすることにより行われる。[13-16]



KEGG [Help](#) [» Japanese](#)

KEGG Home
[Release notes](#)
[Current statistics](#)
[Plea from KEGG](#)

KEGG Database
[KEGG overview](#)
[Searching KEGG](#)
[KEGG mapping](#)
[Color codes](#)

KEGG Objects
[Pathway maps](#)
[Brite hierarchies](#)

KEGG Software
[KegTools](#)
[KEGG API](#)
[KGML](#)

KEGG FTP
[Subscription](#)

[GenomeNet](#)
[DBGET/LinkDB](#)
[Feedback](#)

[Kanehisa Labs](#)

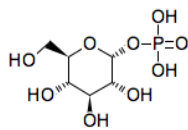
KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (See [Release notes](#) for new and updated features).

- Main entry point to the KEGG web service**
[KEGG2](#) [KEGG Table of Contents](#) [Update notes](#)
- Data-oriented entry points**
 - [KEGG PATHWAY](#) [KEGG pathway maps](#) [[Pathway list](#)]
 - [KEGG BRITE](#) [BRITE functional hierarchies](#) [[Brite list](#)]
 - [KEGG MODULE](#) [KEGG modules](#) [[Module list](#)]
 - [KEGG DISEASE](#) [Human diseases](#) [[Cancer](#) | [Infectious disease](#)]
 - [KEGG DRUG](#) [Drugs](#) [[ATC drug classification](#)]
 - [KEGG ORTHOLOGY](#) [Ortholog groups](#) [[KO system](#)]
 - [KEGG GENOME](#) [Genomes](#) [[KEGG organisms](#)]
 - [KEGG GENES](#) [Genes and proteins](#) [Release history](#)
 - [KEGG LIGAND](#) [Chemical information](#) [[Reaction modules](#)]
- Entry point for wider society**
[KEGG MEDICUS](#) [Health-related information resource](#)
- Organism-specific entry points**
[KEGG Organisms](#) [Enter org code\(s\)](#) [hsa](#) [hsa eco](#)
- Analysis tools**
 - [KEGG Mapper](#) [KEGG PATHWAY/BRITE/MODULE mapping tools](#)
 - [KEGG Atlas](#) [Navigation tool to explore KEGG global maps](#)
 - [KAAS](#) [KEGG automatic annotation server](#)
 - [BLAST/FASTA](#) [Sequence similarity search](#)
 - [SIMCOMP](#) [Chemical structure similarity search](#)
 - [PathPred](#) [Biodegradation/biosynthesis pathway prediction](#)

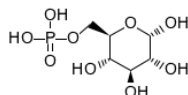
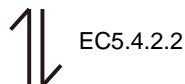
Copyright 1995-2012 Kanehisa Laboratories

図 1-2 KEGG(Kyoto Encyclopedia of Genes and Genomes) データベース



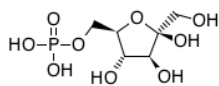
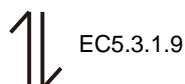
C00103

D-Glucose 1-phosphate



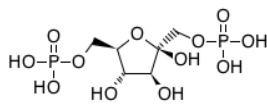
C00668

alpha-D-Glucose 6-phosphate



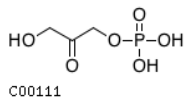
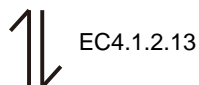
C05345

beta-D-Fructose 6-phosphate



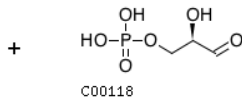
C05378

beta-D-Fructose 1,6-bisphosphate



C00111

Glycerone phosphate



C00118

D-Glyceraldehyde 3-phosphate

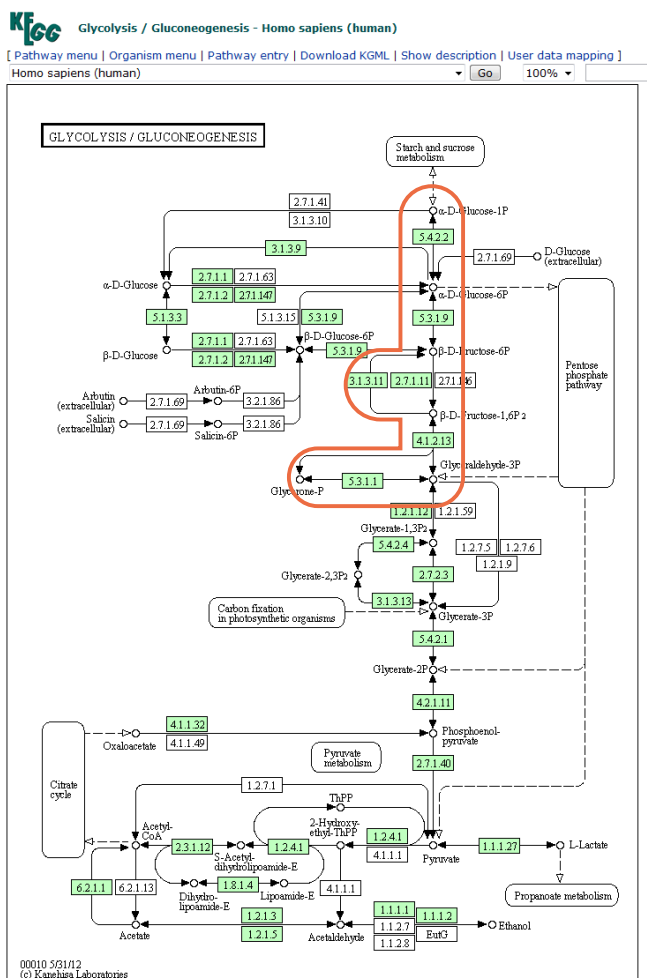


図 1-3 KEGG PATHWAY map

右の赤線で囲われた部分が左の反応経路に相当する。代謝化合物をノード（頂点）、代謝反応をエッジ（辺）として表す。

1.6 代謝ネットワークのモジュール

近年、ゲノム配列解読手法の発展によりゲノム情報を比較的安価に解読できるようになったことにより、計算機によってその生物種の代謝パスウェイを予測する genome-driven な代謝パスウェイ再構築手法は、様々な研究で用いられている。このことにより、新規生物種の生合成能や分解能を予測できるようになっただけでなく、代謝酵素の系統関係を網羅的に調べることで、代謝の進化についても解析が可能になりつつある。しかしながら、広範な代謝ネットワーク中のどの経路が存在すればどのような化合物合成能や分解能と結びつくのかという、代謝の機能単位が明確になっていないことにより、計算機によるゲノムからの機能推定や表現型の進化の解明は未だ困難となっている。

ゲノム解読による代謝の再構築技術が広く使われるようになるにつれ、代謝ネットワーク中の「機能モジュール」の同定への要求が急速に高まっている。モジュール化は、複雑な問題を概念的に理解しやすい「部分」に分解することであり、機能に関連付けられるモジュールは機能モジュールと呼ばれる。ところが、モジュールの定義はそれを提唱した研究者によりそれぞれ異なる。

情報学の一分野であるネットワーク解析の領域では、ネットワークが高い頻度で繋がりあったノードからなるサブネットワークに分割されるとき、それらはモジュール(図 1-4(a))[17]と呼ばれる。これらのノードの集合はネットワーク中で一体となって振る舞い、一つの単位とみなされる。代謝ネットワークの解析においては、代謝の挙動の予測や代謝系の頑健性（ロバストネス）を測るこ

とを目的とした数理モデルの開発に、代謝のモジュール情報が用いられている。ここでのモジュールとは、代謝ネットワークを機能的関連性の高いノードの集合に分割したときに得られるサブパスウェイのことであり、モジュールの中のノードが高頻度で繋がりがあっているとは限らない(図 1-4(b)[18])。むしろ、代謝パスウェイが分岐していないときそれらは機能的に関連しているとみなされるため、ほとんどの反応は前後の反応のみとしか繋がっていない。このように、ネットワーク解析におけるモジュールと代謝ネットワークのモジュールには定義の違いが存在する(図 1-4)が、機能的単位であるという点では一致している。

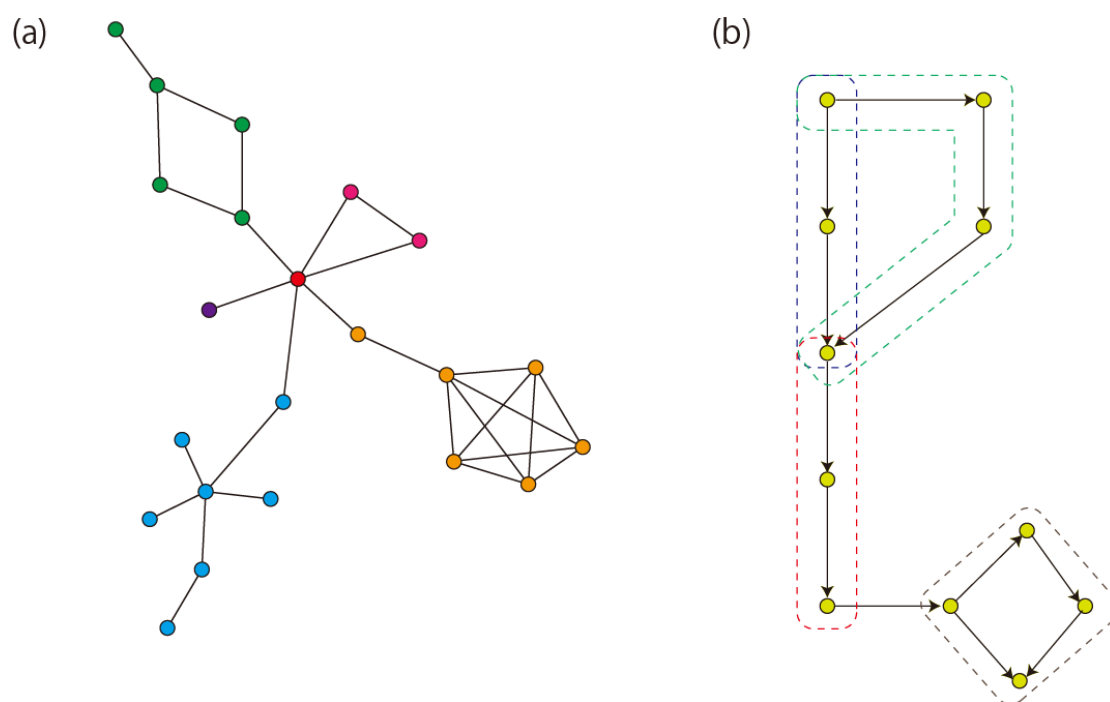


図 1-4 ネットワークモジュール

(a) トポロジカルモジュール (b) 機能関連モジュール

代謝ネットワークのモジュールは、代謝パスウェイ中の機能の定義により、様々な階層のモジュールを考えることができる。例えば、複合体を形成する酵素のサブユニットとなるタンパク質群は物理的相互作用によって定義された機能モジュールと考えることができるし、複数の酵素が関わる反応群や、複数の反応から成る代謝系のサブカテゴリも、機能モジュールと考えることができる。これらの階層的モジュールは、代謝ネットワークを全体から部分へと階層的に定義することで、下位階層の機能の詳細を知らなくても上位階層の機能を理解できるようネットワークを簡略化することができ、複雑な代謝ネットワークの体系的理解を助ける。

先に述べたとおり、代謝ネットワークは遺伝的ネットワークであると同時に、化学的ネットワークでもある。化学的ネットワークの進化には、遺伝的な制限だけでなく、有機反応論に基づいた化学反応性という制限が存在すると考えられる。この制限が存在することは、これまで蓄積された代謝ネットワーク中に、反応を共有していながら遺伝的類縁関係の見られない連続反応が報告されていることから推測される。

これらの保存された連続反応は、生物の代謝能に対して制限を与える有機化学的禁制律に従って、酵素が収斂進化してきた結果であると考えられる。重要なことに、このような禁制律は、好気性・嫌気性や酸性度など生物の生息環境に依存し、より表現型に直結する機能を反映している。これらのことから、我々は代謝ネットワークの階層的理解には、化学反応によって機能を定義されたモジュール(化学反応モジュール)の階層が必要であると考えた。

1.7 パスウェイアライメント

代謝ネットワークから保存された反応列を見つけ出すためには、異なる基質に対して起こる反応間の類似性を定義し、経路の類似性を検出するパスウェイアライメントの手法が必要となる。経路の類似性は、経路を構成する反応の類似性の総和で定義され、反応間の類似性をどのように定義するかが要となる。パスウェイアライメント問題に対しては、これまで様々な手法が開発されてきた。酵素間の類似性には、従来法ではアミノ酸配列類似性や上述のスーパーファミリーを用い、オーソログ酵素遺伝子をパスウェイ中の該当する反応に当てはめる。この手法では、酵素タンパク質のアミノ酸配列が決定された代謝反応のみが対象となる。この方法の限界としては、既に述べたように、配列類似性が必ずしも反応類似性を意味しないことが挙げられる。Tohsato らは EC 番号の類似性を酵素間の類似性と定義し、代謝ネットワークから類似した反応ステップを見つけ出す手法を提案した[19]。Pinter らは Tohsato の EC 番号による定義を踏襲し、異なる生物種や多様化した経路の間で保存された経路を検出する手法を開発した[20]。Wernicke らは Pinter の代謝パスウェイアライメント手法をもとに、より高速なアルゴリズムを提供し[21]、Ay らは EC 番号類似性と代謝化合物類似性とを結びつけ、代謝ネットワークのトポロジーの類似性をパスウェイアライメントに利用した[22]。

上記の、オーソログ酵素遺伝子を該当する反応に割り当てる方法は、同一の酵素により触媒される同一の反応からなるパスウェイしか扱うことができない。

EC 番号類似性を用いた手法では同一の反応だけでなく、EC 番号の階層に基づき類似した反応も考慮することが可能である。しかしながら、EC 番号は実験的に活性が確認された酵素とその反応に対し酵素委員会の承認を経て与えられた、酵素の分類であり、EC 番号の類似性は必ずしも反応の類似性を保証しない。また、分類階層の深さと反応の類似度の関連性の高さは異なる分類群間で一致しておらず、類似性をスコアで定義することが難しい。例えば、一部の EC 階層においては、主産物の有機反応の分類よりも補酵素の一致などが優先され、必ずしも有機反応の類似性が反映されていない例が散見される(図 1-5)。さらに、EC 番号は酵素委員会により承認の得られた酵素だけではなく、ゲノム中の遺伝子に対し配列類似性に基づいてアノテーション（注釈付け）されている例が多いため、純粹に反応の類似性によって分類されたものではない。実際に多くの解析において、配列類似性に基づく方法と大きな差が無いことが示されている。したがって EC 番号の類似性は、酵素触媒反応の化学的多様性の比較のようなシステムティックな解析に適した基準であるとは言えない。

(a) 異なる EC sub-sub classes に分類されたヒドロキシラーゼ反応

EC: 1.14.13	<p>R05736 (EC: 1.14.13.72)</p>	
EC: 1.14.14	<p>R01840 (EC: 1.14.14.1)</p>	
EC: 1.14.15	<p>R03262 (EC: 1.14.15.5)</p>	

(b) 同一の EC sub-subclass に分類されたヒドロキシラーゼ反応

EC: 1.14.13	<p>R00815 (EC: 1.14.13.7)</p>	
	<p>R00448 (EC: 1.14.13.59)</p>	

図 1-5 EC 階層分類の問題点の例

ヒドロキシラーゼの EC 階層分類。(a)基質と産物の有機反応は類似しているにも拘らず、補酵素の違いによって異なる EC sub-sub classes に分類されたヒドロキシラーゼ反応 (b) 反応中心となる原子種が異なるにも拘らず、補酵素が同一であるために同一の EC sub-sub classes に分類されたヒドロキシラーゼ反応 Reprinted and modified with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright 2013 American Chemical Society.

1.8 本研究の目的

本研究では、代謝反応の化学に焦点を当てた。我々はまず、「ゲノム上で連続して存在する遺伝子クラスターによって明らかにされた遺伝的機能単位は、有機反応の化学的機能単位を反映している」という作業仮説を立てた。そこで我々は酵素タンパク質や酵素遺伝子の配列情報を用いることなしに、化学構造変化だけに基づいて反応モジュールを抽出する手法を開発した。

前項に述べたこれまでのパスウェイアライメントの問題点を考慮し、我々は、酵素タンパク質や酵素遺伝子の配列情報を用いることなしに、化学構造変化だけに基づいて反応モジュールを抽出する手法を開発する必要があると考えた。そこで本著ではまず、パスウェイアライメントのための新規の類似性指標を提案する。本指標は、代謝パスウェイ中の化学構造変換パターンの類似性に基づき、タンパク質の配列情報や EC 番号を一切用いない化学的類似性指標である。本手法では、反応類似性の定義に基質と産物の化合物ペアから抽出された構造上の局所的変化を利用した。データセットは KEGG データベース[12]中の KEGG RPAIR (reactant pair)データベース及び KEGG RCLASS(reaction class)データベースから取得した。KEGG RPAIR は既知の酵素触媒反応の基質・産物ペアとその化学構造変換パターンを集めたものである。KEGG RCLASS は、代謝反応の主な基質と産物のペアに相当する”main”というタイプ(後述)に分類される化合物ペアについて、同一の構造変化パターンを有するものを同一の反応クラスとしてまとめたデータベースである。

上記の化学反応類似性指標を用いて、代謝ネットワークデータベースから保存された連続反応の抽出を行った。代謝パスウェイ上の連続反応のパターンが異なる経路に保存されて出現するとき、本研究ではこの連続反応パターンを反応モジュールと定義した。この反応モジュールを、**RCLASS** エントリ間の類似性スコアを用いて、**KEGG** 代謝パスウェイから体系的に抽出した。得られた反応モジュールを、ゲノム上で連続して存在する遺伝子クラスターによって明らかにされた遺伝的機能単位を集めた、**KEGG MODULE** データベース中のパスウェイモジュールと比較した。

第2章 データセット

2.1 代謝パスウェイデータベース

本解析ではデータソースとして KEGG データベース (<http://www.kegg.jp/>) release 62.0+ (May 24, 2012)を使用した。KEGG PATHWAY データベースは、文献から収集された分子間相互作用や反応ネットワークに関する知識のコレクションであり、集められた知識はマップ上に表記することにより可視化されている。我々は KEGG PATHWAY データベースの代謝のセクションに蓄積されているパスウェイ情報を用いた。典型的な KEGG 代謝パスウェイマップは、低分子化合物が酵素により触媒される反応によってどのように変換されるかを描写している。KEGG PATHWAY では代謝化合物をノード (頂点)、代謝反応をエッジ (辺) として表し(図 1-3)、代謝反応に KEGG ORTHOLOG データベース中のオーソログ情報との関連付けがなされている。通常、KEGG PATHWAY データベース中のマップは種特異的ではなく、異なる生物種からの情報を統合したリファレンスパスウェイと呼ばれる一つのマップに表わされている。このリファレンスパスウェイマップ上の反応に関連付けられたオーソログ遺伝子から、目的のゲノムに含まれる遺伝子を探し出すことで、種特異的なパスウェイを生成し利用できる。つまり、遺伝子のネットワークと反応のネットワークという代謝系の二つの側面は、KEGG REACTION データベースの酵素反応のエントリに対応する KEGG リファレンスパスウェイマップ上の各々の反応が、KEGG ORTHOLOGY データベースに蓄積された酵素のオーソログ遺伝子エントリと対応付けられる

ことによって表現されている。KEGG は XML (Extensible Markup Language) 形式でもデータを提供しており、パスウェイマップの XML ファイルは隣り合った酵素の関係についての情報を含んでいる。本研究ではリファレンスパスウェイの XML ファイルから得た情報を元に、代謝パスウェイからの連続した反応ステップの抽出を行った。

2.2 反応ペア

KEGG REACTION データベースは IUBMB(The International Union of Biochemistry and Molecular Biology)の承認した酵素反応[2]および KEGG が独自に文献より収集した既知の全酵素反応を蓄積する。本研究で用いたバージョンの KEGG リリースには 8,990 の反応が蓄積されている。KEGG 代謝パスウェイ上に出現する反応に限れば、6,238 の反応が蓄積される。通常、酵素反応は酵素反応式という形で表され、一つの反応に複数の基質と生成物が関わっている例も多い。KEGG RPAIR データベースでは酵素反応中の水素以外の原子の流れに基づいて、基質と生成物の一対一関係を反応ペアとして定義している。つまり、一つの酵素反応式は、基質・生成物をノード（頂点）とし、それらを結ぶ複数の反応ペアをエッジ（辺）とする反応ペアグラフ(図 2-1)として表現されていると言える。本研究で用いたバージョンの KEGG RPAIR データベースには、13,448 の反応ペアが蓄積されている。各々の反応ペアの構造変換パターンは、化学構造比較プログラム SIMCOMP[23]を用い基質・生成物間の原子-原子マッピングを

自動的に計算した後、専門家による修正を加えた結果を元に、RDM 表記法（後述）によって記述されている。

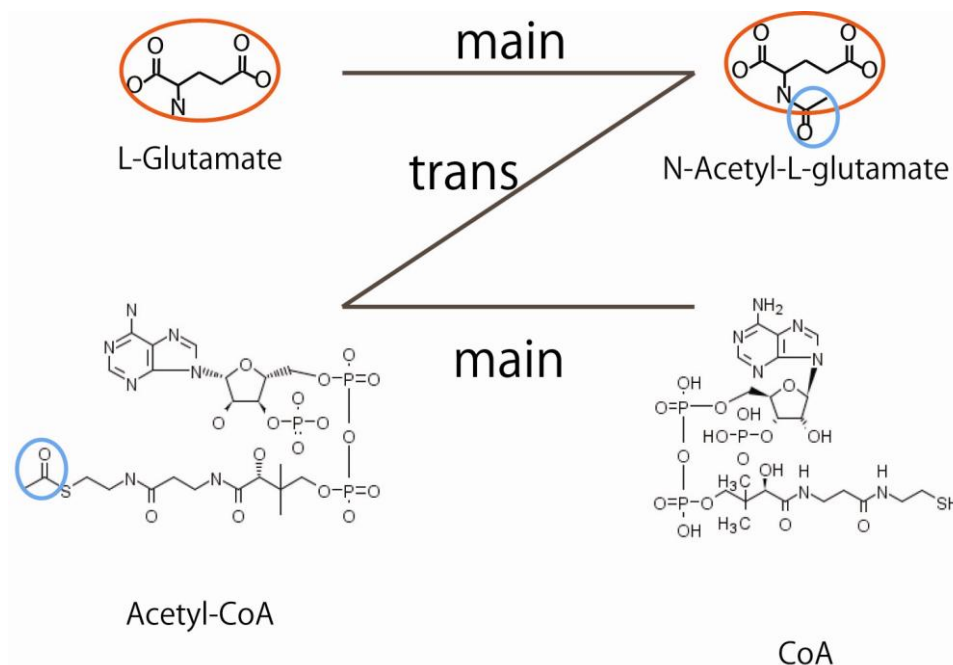


図 2-1 反応ペアグラフ

全ての反応ペアは、反応ペアグラフのトポロジー（接続関係）や原子種の情報に基づき *main*、*trans*、*cofac*、*leave* 及び *ligase* の 5 種類のタイプにクラス分けされている。原子の主な流れを示す反応ペアグラフ（図 2-1）中で、少なくとも一つのペアは *main* に分類される。*main* に分類される反応ペアは、例えば酸化還元反応から NAD⁺などの補酵素を除いた基質・生成物のペアなど、その反応の主たる反応であると思われる反応ペアであり、KEGG 代謝パスウェイ上に現れる化合物ペアと一致する場合が多い。酸化還元反応の補酵素のペアは *cofac* に分類され、合成反応に関わる ATP などの三リン酸は *ligase* というタイプが与えられている。*trans* に分類される反応ペアは転移酵素によって転移される官能

基（例えばアシル転移酵素におけるアシル基など）を表現するペアであることが多いが、厳密な定義としては「反応ペアグラフ中で二つの化合物（ノード）がそれぞれ二つ以上のエッジ（辺）を持つような反応ペア」である。ただし、一つの酵素反応式から反応ペアグラフを形成した場合に *trans* に分類されていた反応ペアであっても、代謝パスウェイ中の化合物の流れの中で一つのステップを形成するとき、*main* に分類し直されることがあることに注意されたい。本研究では解析の対象を *main* に分類された反応ペアに限定した。

2.3 KEGG atom type

KEGG 中の低分子化合物は KEGG Chemical Function (KCF)形式のファイルでグラフ構造として記述されている[23]。KCF 形式では、同じ原子種（炭素原子、窒素原子など）であっても異なる官能基中の原子は異なる KEGG atom type として区別されている。（表 2-1）KEGG atom type は原則三つの文字から成る。一文字目の英大文字は原子種を、二文字目の数字は各々の原子種に対し定義された原子間結合の種類を、三文字目の英小文字は置換基の数など定義されたトポロジー情報を示す。例えばアルデヒドの炭素原子は C4a、第一級アミンの窒素原子は N1a のように表される。生体低分子中の重要な官能基を区別できるよう、炭素 23 種類、窒素 16 種類、酸素 18 種類、硫黄 7 種類、リン 2 種類、及びハロゲン原子とその他の原子に対して 1 種類ずつの、合計 68 の atom type がこれまでに定義されている。

表 2-1 KEGG atom type

Reprinted with permission from [J Chem Inf Model.](#), (2013)53, 613-622. Copyright(2013) American Chemical Society.

Atom	Atom type	Functional group	Description	Atom	Atom type	Functional group	Description
C	C1a	Alkane	R-CH3	O	O1a	Hydroxy	R-OH
	C1b		R-CH2-R		O1b		N-OH
	C1c		R-CH(-R)-R		O1c		P-OH
	C1d		R-C(-R)2-R		O1d		S-OH
	C1x	Cyclic alkane	ring-CH2-ring		O2a	Ether	R-O-R
	C1y		ring-CH(-R)-ring		O2b		P-O-R
	C1z		ring-CH(-R)2-ring		O2c		P-O-P
	C2a	Alkene	R=CH2		O2x		ring-O-ring
	C2b		R=CH-R		O3a	Oxo	N=O
	C2c		R=C(-R)2		O3b		P=O
	C2x	Cyclic alkene	ring-CH=ring		O3c		S=O
	C2y		ring-C(-R)=ring or ring-C(=R)-ring		O4a	Aldehyde	R-CH=O
	C3a	Alkyne	R≡CH		O5a	Ketone	R-C(=O)-R
	C3b		R≡C-R		O5x		ring-C(=O)-ring
	C4a	Aldehyde	R-CH=O		O6a	Carboxylic acid	R-C(=O)-OH
	C5a	Ketone	R-C(=O)-R		O7a	Ester	R-C(=O)-O-R
	C5x	Cyclic ketone	ring-C(=O)-ring		O7x		ring-C(=O)-O-ring
	C6a	Carboxylic acid	R-C(=O)-OH		O0	Undefined O	
	C7a	Carboxylic ester	R-C(=O)-O-R	S	S1a	Thiol	R-SH
	C7x		ring-C(=O)-O-ring		S2a	Thioether	R-S-R
	C8x	Aromatic ring	ring-CH=ring		S2x		ring-S-ring
	C8y		ring-C(-R)=ring		S3a	Disulfide	R-S-S-R
	C0	Undefined C			S3x		ring-S-S-ring
					S4a	Sulfate	R-SO3
N	N1a	Amine	R-NH2	P	S0	Undefined S	
	N1b		R-NH-R		P1a	Attached to other elements	P-R
	N1c		R-N(-R)2	Other	P1b	Attached to oxygen	P-O
	N1d		R-N(-R)3+		X	Halogens	F, Cl, Br, I
	N1x	Cyclic amine	ring-NH-ring	Z		Others	
	N1y		ring-N(-R)-ring				
	N2a	Imine	R=N-H				
	N2b		R=N-R				
	N2x	Cyclic imine	ring-N=ring				
	N2y		ring-N(-R)+=ring				
	N3a	Cyan	R≡N				
	N4x	Aromatic ring	ring-NH-ring				
	N4y		ring-N(-R)-ring				
	N5x		ring-N=ring				
	N5y		ring-N(-R)+=ring				
	N0	Undefined N					

2.4 RDM 化学構造変換表記法

各々の反応ペアは、まず基質と産物の化学構造を重ね合わせることで、一致部分構造 (M-substructure) と相違部分構造 (D-substructure) に分けられる。一致部分構造に位置する原子のうち相違部分構造に隣接する原子を反応中心原子 (R 原子) とし、R 原子と共有結合する原子のうち、相違部分構造側の原子を D 原子、一致部分構造側の原子を M 原子とする (図 2-2)。反応の前後におけるこれら 3 つの原子 (R 原子、D 原子、M 原子) の KEGG atom の変化を示すことで化学変換パターンを表現したものを、RDM 表記法と呼ぶ。反応ペア A,B について RDM 表記法は以下のようになる：

$$\text{RDM}(A, B) = \text{RA-RB} : \text{DA-DB} : \text{MA-MB}$$

例えば、第一級アミンで起こる典型的なアシル転移反応 (図 2-2) は以下のように表される：

$$\text{RDM} = \text{N1a-N1b} : *-\text{C5a:C1b-C1b}$$

上記の表記法中では、*は置換反応における、KEGG atom types で定義されていない水素原子に対応する。*が表記中に現れることは、水脱離反応を表す RDM (例：C1c-C1b:O1a-*:C1b+C1b-C1b+C4a) など、相違部分構造中に原子が存在しないことを意味する場合があることに注意されたい。また、一つの RDM 表記法は通常、反応内で起こった一つの化学結合変化を表す。一つの反応が二つ以上の化学結合を生成したり分解したりする場合には、二つ以上の RDM 表記が必要になる。[24]

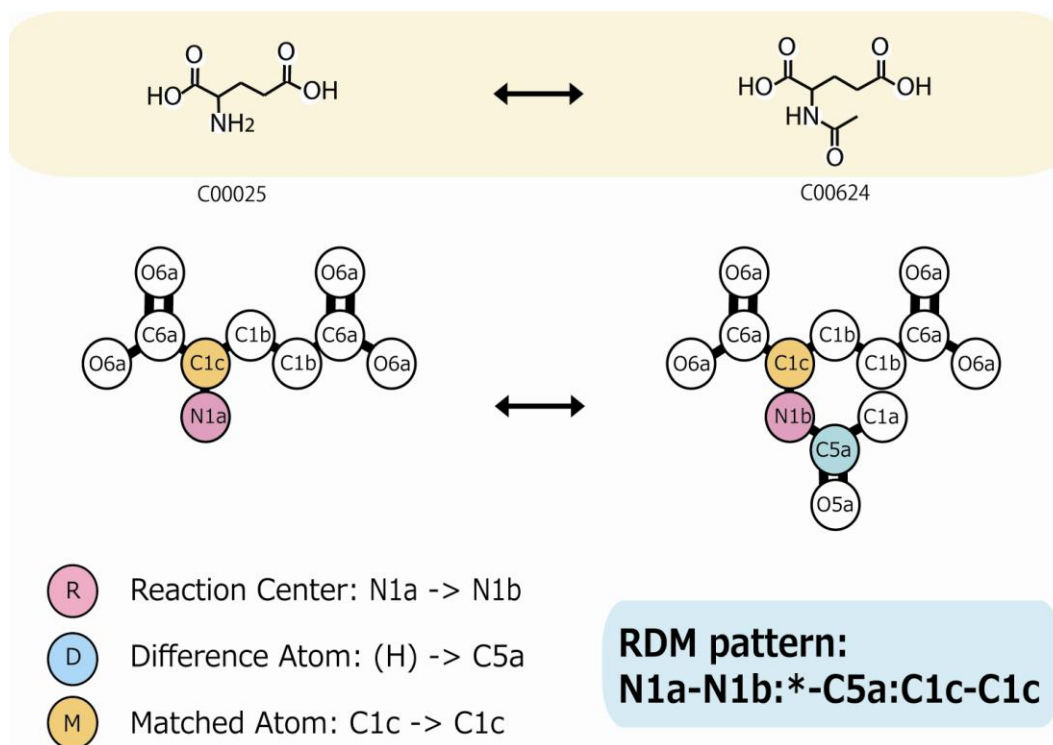


図 2-2 RDM

化学構造比較プログラム SIMCOMP[23]は与えられた2つの分子構造について、それらの間の最大共通部分構造を計算する。この計算は KEGG atom type をノードとしたグラフ構造アライメント手法により行われる。得られた共通部分と非共通部分構造の境界に位置するノードは反応中心原子(Reaction center atom)と定義され、反応中心原子に隣接する非共通部分側のノードは Difference atom、共通部分側のノードは Matched atom と定義される。図中のペアに対する Reaction center atom は N1a(C00025) \rightleftharpoons N1b(C00624)となる。Reaction center atom、Difference atom、Matched atom の3種類のノードを両構造から抽出し組み合わせたものを RDM パターンと呼び、本研究ではこれを基質・産物間の構造変換パターンと定義した。KCF 形式では水素原子は省略されるため、C00025 側の Difference atom は*で表記される。

酵素反応による化学構造変換のパターンを得る最初のステップは、反応ペアの原子-原子マッピングに基づき反応中心原子を決定することである。SIMCOMP プログラム[23]によって生成された化学構造アライメント結果を基に、反応ペア間の一致部分構造 (M-substructure) を決定する。残りの相違部分構造 (D-substructures) は反応中に付加あるいは脱離した原子グループであるとみなされる。上では説明の簡略化のため、M-substructure の原子のうち D-substructure と接する原子が反応中心原子(R 原子)であると説明したが、厳密には R 原子の決定にはより複雑なルールが適用される。(1) D-substructure に相当する原子 (または原子群) が反応ペア中に存在する場合は M-substructure 中の原子のうち D-substructure に結合する原子を R 原子とする。(2) D-substructure に相当する原子が存在せず、少なくとも一つの原子の酸化数が変化している場合には、酸化還元反応 (又は水素付加/脱離) が起こっているとみなし、酸化数の変化した原子を R 原子とする。(3) D-substructure に相当する原子が存在せず、酸化数の変化した原子も存在しない場合には、立体異性 (R/S 異性) のキラル中心または幾何異性 (E/Z 異性) した二重結合原子を R 原子とする。

反応中心原子(R 原子)が決定したら、それに隣接する原子が定義される。R 原子に隣接する原子のうち、D-substructure 中で R 原子に最も近傍に存在する原子が D 原子、M-substructure 中のそれに M 原子と見なされる。反応ペアの化学構造変換は R, D 及び M 原子の KEGG atom 変化を示す RDM 化学構造変換表記法により表現される。

2.5 反応クラス

KEGG RCLASS データベースは、KEGG 代謝パスウェイマップ上に現れる反応の化学構造変換パターンをクラス分けしたものである。ここでいう化学構造変換パターンとは、一つの “*main*” 反応ペア中に含まれる RDM であり、ペアによっては複数の RDM から構成されていることがある。データベース中のエントリーは RC 番号 (“RC” というアルファベットとそれに続く五桁の数字) という ID 番号で識別される。言い換えれば、各々の RCLASS エントリーは KEGG 代謝パスウェイマップ上に現れる基質-生成物ペアのうち、化学変換パターンが同一であるものの集合を示すことになる。RCLASS エントリーは RPAIR データベースから自動的に生成され、化学構造変換パターンを表す図やその他の情報は手作業で付加されている。本研究で用いた RCLASS データベース中の RCLASS エントリー数は 2,481 である。

第3章 結果1：代謝反応クラス類似性比較手法の開発

3.1 フィンガープリント表記

RDM 化学変換パターン及びそれに対応する RCLASS エントリは、本研究の目的に対しては分類が細かすぎるため、本研究では RCLASS エントリ間の類似性スコアリング手法を導入した。そのためにまず、KEGG atom type を 12 のキーを用いたフィンガープリント(表 3-1)で表記した。

これらのキーは炭素原子、カルボニル基の炭素原子、 π 結合に参加する炭素原子、酸素原子、不対電子を持つ酸素原子、窒素原子、リン原子、硫黄原子、ハロゲン、その他の金属原子及び芳香族環原子や環構造中の原子の存在の有無を示す。KEGG atom type とキーの対応を表 3-2 に示す。次に RDM 表記を 72 ビットのフィンガープリント表記に変換した(図 3-1)。例えば、メチル(C1a)、メチレン(C1b)及びその他の sp^3 炭素原子 (C1c、C1d) は異なる KEGG atom types が与えられているが、フィンガープリント表記では同一のものとした。

表 3-1 フィンガープリントキー

Key	Feature	Example
C	Carbon atom	Alkane, Alkene, Alkyne
K	Carbonyl	Aldehyde, Ketone, Ester
E	C with p electron	Alkene, Alkyne
O	Oxygen atom	Hydroxy, Ether, Oxidation
U	O with unpaired electron	Oxo, Aldehyde, Ketone, Ester
N	Nitrogen atom	Amine, imine, Cyan
P	Phosphorus atom	Phosphate
S	Sulfur atom	Thiol, sulfate
X	Halogen	F, Cl, Br, I
Z	Others	Mn, Zn
A	Aromatic Ring	Aromatic
R	Ring	Cyclic

		R		D		M	
		R1x	R1y	D1x	D1y	M1x	M1y
RDM1	C5a	-	C1d	*	-	C5a	
F1	C K E O U	C K E O U	C K E O U	C K E O U	C K E O U	C K E O U	C K E O U
	1 1 0 0 0	1 0 0 0 0	0 0 0 0 0	1 1 0 0 0	1 1 0 1 1	1 1 0 1 0	

		R		D		M	
		R2x	R2y	D2x	D2y	M2x	M2y
RDM2	C5a	-	C1d	*	-	C5a	
F2	C K E O U	C K E O U	C K E O U	C K E O U	C K E O U	C K E O U	C K E O U
	1 1 0 0 0	1 0 0 0 0	0 0 0 0 0	1 1 0 0 0	1 1 0 1 1	1 1 0 1 0	

図 3-1 RDM 表記のフィンガープリント表記への変換

ピルビン酸とアセト乳酸（上）及びオキソブタン酸と 2-アセト-2-ヒドロキシブタン酸（下）の反応ペアに対応する二つの RCLASS エントリ：RC00106 及び RC01192 を示す。これらの RCLASS エントリに対する RDM 表記は M 原子(円で囲んだ)に相違があるものの、フィンガープリント表記では同一として扱われる。図では簡単のため、12 のキーのうちこれらの反応ペアに参与する 5 つのキーのみを表示している。Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright 2013 American Chemical Society.

3.2 RC エントリの類似性スコア

比較する二つの RCLASS エントリがそれぞれ一つずつの RDM しか持たない場合には、エントリ間の類似性スコアは二つの RDM 間の類似性スコアに等しい。比較対象の RDM はフィンガープリントキー(表 3-1、図 3-1)で定義された二値特徴ベクトルのセットへと変換される。

$$F_1 = RDM_1 = \begin{bmatrix} v_{R1x} & v_{R1y} \\ v_{D1x} & v_{D1y} \\ v_{M1x} & v_{M1y} \end{bmatrix}, F_2 = RDM_2 = \begin{bmatrix} v_{R2x} & v_{R2y} \\ v_{D2x} & v_{D2y} \\ v_{M2x} & v_{M2y} \end{bmatrix}$$

ここで、x と y は反応式中の左辺と右辺の化合物から定義されたベクトルを指す。

二値特徴ベクトル v_1 と v_2 の間の類似性スコアは以下のジャッカード係数で定義した：

$$J(v_1, v_2) = \frac{|v_1 \wedge v_2|}{|v_1| + |v_2| - |v_1 \wedge v_2|}$$

$|v|$: 特徴ベクトル v 中の 1 の数 (キーの存在数)

$|v_1 \wedge v_2|$: 特徴ベクトル v_1 と v_2 の双方に 1 がアサインされているキーの数

RDM_1 と RDM_2 の間の類似性スコア S は以下のように定義した：

$$\begin{aligned} S(RDM_1, RDM_2) &= w_R J_a(v_{R1}, v_{R2}) + w_D J_a(v_{D1}, v_{D2}) + w_M J_a(v_{M1}, v_{M2}) \\ &= S(RC_1, RC_2) \end{aligned}$$

ただし

$$J_a(v_1, v_2) = \frac{J(v_{1x}, v_{2x}) + J(v_{1y}, v_{2y})}{2}$$

また

$$\left\{ \begin{array}{ll} \text{(i)} & \text{if } v_{D1x} = v_{D2x} = v_{D1y} = v_{D2y} = \emptyset, \\ & w_R = 0.5, \quad w_D = 0, \quad w_M = 0.5 \\ \text{(ii)} & \text{if } (v_{D1x} \neq \emptyset \text{ or } v_{D1y} \neq \emptyset) \text{ and } (v_{D2x} \neq \emptyset \text{ or } v_{D2y} \neq \emptyset) \\ & w_R = 0.4, \quad w_D = 0.4, \quad w_M = 0.2 \\ \text{(iii)} & \text{if } \{ (v_{D1x} = v_{D1y} = \emptyset) \text{ and } (v_{D2x} \neq \emptyset \text{ or } v_{D2y} \neq \emptyset) \} \text{ or } \{ (v_{D2x} = v_{D2y} = \emptyset) \text{ and } (v_{D1x} \neq \emptyset \text{ or } v_{D1y} \neq \emptyset) \} \\ & w_R = 0, \quad w_D = 0, \quad w_M = 0 \end{array} \right. \quad \text{ここで } \emptyset \text{ は zero vector}$$

フィンガープリント v は R , D 及び M 毎にそれぞれ独立して比較される。また、ジャッカード係数 J_a は例えば D 原子が存在するかどうかなどによって重みづけした。

少なくとも一つの **RCLASS** エントリが複数の **RDM** を持つ場合には、類似性スコアは以下のように定義した。(1) まず全ての可能な **RDM** の組み合わせ (**RDM**

ペア) について類似性スコアを計算し、スコアの大きい順に並べる。このとき、スコア 0.7 未満のペアは捨てる。(2) RDM ペアのうち、どちらか片方の RDM でも既に選択済みであれば、そのペアを捨て、次のペアを見る。(3) RDM ペアのうち、どちらの RDM とも未選択であれば、そのペアを選択し、次のペアを見る。(4) プロセス 2 と 3 を、RDM ペアが無くなるまで繰り返す。(5) 以上のようにして、選択された (マッチした) 全ての RDM ペアについて類似性スコアの平均値を計算し、マッチしなかった RDM に対するペナルティスコアを減じたものを、その RC1 と RC2 の類似性スコアと定義する。

$$S(RC_1, RC_2) = \frac{\sum_{i=0}^{N_m} S(RDM_1, RDM_2)}{N_m} - w_p \cdot \frac{N_u}{N_m + N_u}$$

w_p : アンマッチペナルティ = 0.5

N_m : マッチした RDM の数($S(RDM_x, RDM_y) \geq 0.7$) RDMs

N_u : マッチしなかった RDM の数

表 3-2 KEGG atom type と フィンガープリントキーの対応

Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright(2013) American Chemical Society.

Atom	Atom type	Fingerprint keys	Atom	Atom type	Fingerprint keys
C	C1a	C	O	O1a	O
	C1b	C		O1b	O
	C1c	C		O1c	O
	C1d	C		O1d	O
	C1x	C R		O2a	O
	C1y	C R		O2b	O
	C1z	C R		O2c	O
	C2a	C E		O2x	O
	C2b	C E		O3a	O U
	C2c	C E		O3b	O U
	C2x	C E R		O3c	O U
	C2y	C E R		O4a	O U
	C3a	C E		O5a	O U
	C3b	C E		O5x	O U
	C4a	C K		O6a	O U
	C5a	C K		O7a	O U
	C5x	C K R		O7x	O U
	C6a	C K		O0	O
	C7a	C K	S	S1a	S
	C7x	C K R		S2a	S
N	C8x	C A		S2x	S
	C8y	C A		S3a	S
	C0	C		S3x	S
	N1a	N		S4a	S
	N1b	N		S0	S
	N1c	N	P	P1a	P
	N1d	N		P1b	P
	N1x	N R	Other	X	X
	N1y	N R		Z	Z
	N2a	N			
	N2b	N			
	N2x	N R			
	N2y	N R			
	N3a	N			
	N4x	N A			
	N4y	N A			
	N5x	N A			
	N5y	N A			
	N0	N			

3.3 RCLASS エントリの類似性によるグループ化

本解析では RCLASS エントリの類似性スコアの閾値を 1.0 としてグループ化した (他の閾値は他の解析で利用した)。この閾値の意味するところは、RCLASS エントリをグループ分けするために、RDM 表記をフィンガープリント表記に変換し、その表記の完全一致するものを同じグループと見なしたということになる。グループ化の結果、2,481 の RCLASS エントリは 376 の類似グループ (複数のエントリから成るグループ) と、1,190 のシングルトン (単一のエントリから成る、他のどのエントリともグループ化されなかったもの) に分けられた。

3.4 RCLASS 保存配列パターンの抽出

上で定義された RCLASS エントリ間の類似性指標を用い、我々は代謝ネットワークからの連続反応パターンの抽出を行った。ここで連続反応パターンとは、異なる代謝パスウェイから抽出された連続反応に共通して出現する RCLASS の保存配列パターンと定義した。連続反応パターンの抽出は以下の方法で行った。

KEGG PATHWAY データベース中の既知の代謝パスウェイを、全ての可能な 2~8 個の線形な連続反応列に分割した。例えば、5 つの連続反応(A-B-C-D-E)からなる線形経路からは、長さ 2 の反応列が 4 つ(A-B, B-C, C-D, D-E)、長さ 3 の反応列が 3 つ(A-B-C, B-C-D, C-D-E)、長さ 4 の反応列が 2 つ(A-B-C-D, B-C-D-E)、及び長さ 5 の反応列が 1 つ(A-B-C-D-E) 得られ、それ以外の長さの反応列は得られない。分岐を含むパスウェイの場合は、全ての組み合わせの線形な連続反応

列へと分割した(図 3-2 (a))。経路が不可逆反応を含む場合には、逆方向の反応を含む連続反応列は除外した。環状経路については、全ての可能な連続反応列を生成したが、同一の反応が二度以上出現するような反応列は除外した。

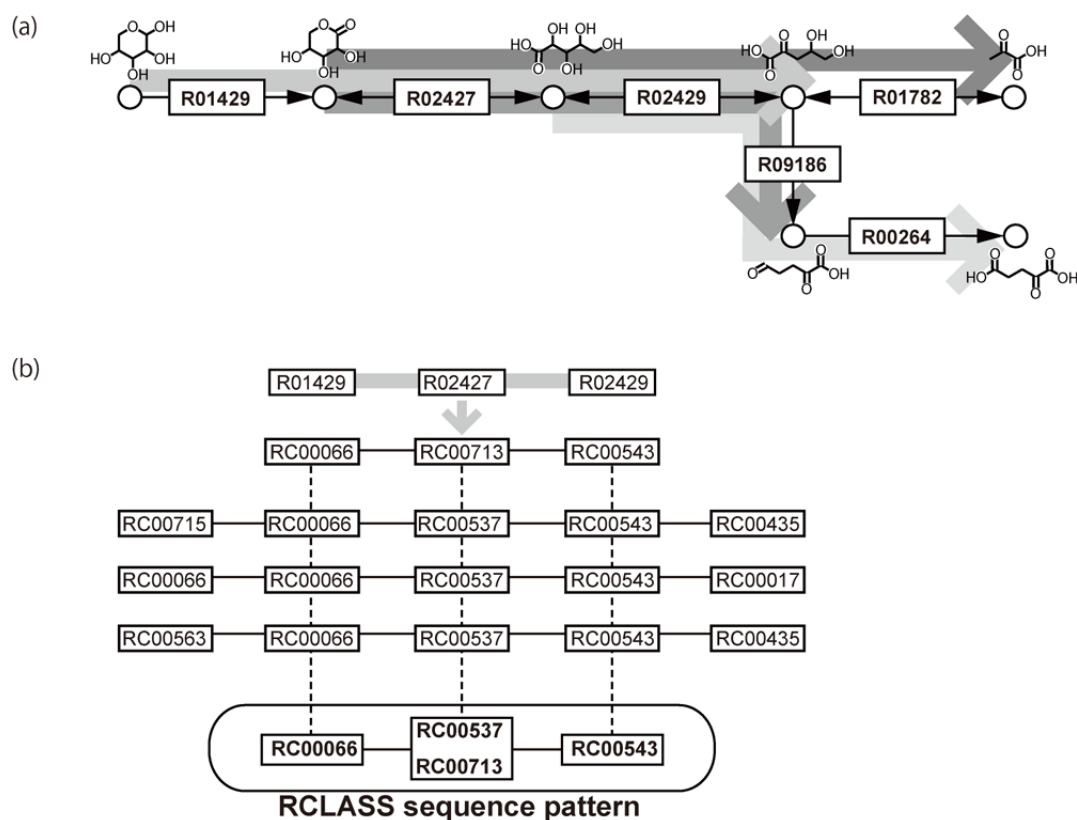


図 3-2 パスウェイからの保存連続反応パターンの抽出

(a) 代謝パスウェイからの長さ 3~8 (図中では 3) の反応列の抽出 (b) 反応 (R 番号) 列の RCLASS (RC 番号) 配列への変換及び共通 RCLASS (RC 番号) 配列パターンの抽出

Reprinted with permission from [J Chem Inf Model.](#), (2013)53, 613-622. Copyright 2013 American Chemical Society.

次に、連続反応列を構成する反応の ID 変換を行った。生成された反応列は、KEGG の酵素反応を意味する R 番号（アルファベット R とそれに続く 5 桁の数字）が 2 ～ 8 個だけ連なった列から成っている。その R 番号を、それが属する RCLASS を意味する RC 番号（アルファベット RC とそれに続く 5 桁の数字）の配列に変換した。こうすることで、連続反応列を構成する個々の反応が違っていても（R 番号が違っていても）、同じ化学構造変換を連続して起こす（同じ RC 番号を持つ）場合には、それらは同一であるとみなされるようにした。また、RC 番号が違っていても、フィンガープリント表記により同じ類似グループに属する場合には類似した化学構造変換であるとみなした。以上のようにして、KEGG 代謝パスウェイ全体からそれぞれの長さについての RCLASS 保存配列パターンを抽出した。

第4章 結果2：化学反応モジュール

4.1 抽出された保存反応列パターン

代謝パスウェイデータベースから抽出した2～8個の長さの連続反応配列をそれぞれの長さごとに独立に計算し、得られた保存配列パターンのうち短いパターンが長いパターンに完全に含まれている場合にはそれらを取り除いた。その結果、3,016の保存配列パターンが得られた。得られた保存配列の長さと数を表4-1に示した。パスウェイを構成する6,238反応中でRCLASSのアサインされている5,805反応のうち、およそ60%が保存された配列パターンに含まれていた。我々は手作業で結果を精査し、生化学的意味付けしたものを反応モジュールと定義した。(表4-2)

4.2 反応モジュールの出現頻度

反応モジュールの分布の傾向として、短いものほど多種類のパスウェイに出現し、長いものほど一部の代謝系に偏って存在していた(図4-1)。最も多くのパスウェイから得られたのはリン酸化の繰り返し反応であった。本研究では最長8反応までのモジュールを抽出したが、8反応のモジュールの多くはリン酸化または配糖化の繰り返しを含むモジュールであった。繰り返し反応を除くと、最も長い反応モジュールは生分解経路に出現するキシレン分解モジュールであった。

表 4-1 KEGG 代謝パスウェイ中に見つかった RCLASS 保存配列パターンの数

Length	# of conserved patterns	# of reactions included	Coverage*
2	928	3,479	0.599
3	770	2,503	0.431
4	534	1,662	0.286
5	338	1,074	0.185
6	218	765	0.132
7	140	527	0.091
8	88	399	0.069
Total	3,016		

* KEGG 代謝パスウェイ上で RC が与えられた反応の総数 5,805 反応に対する割合

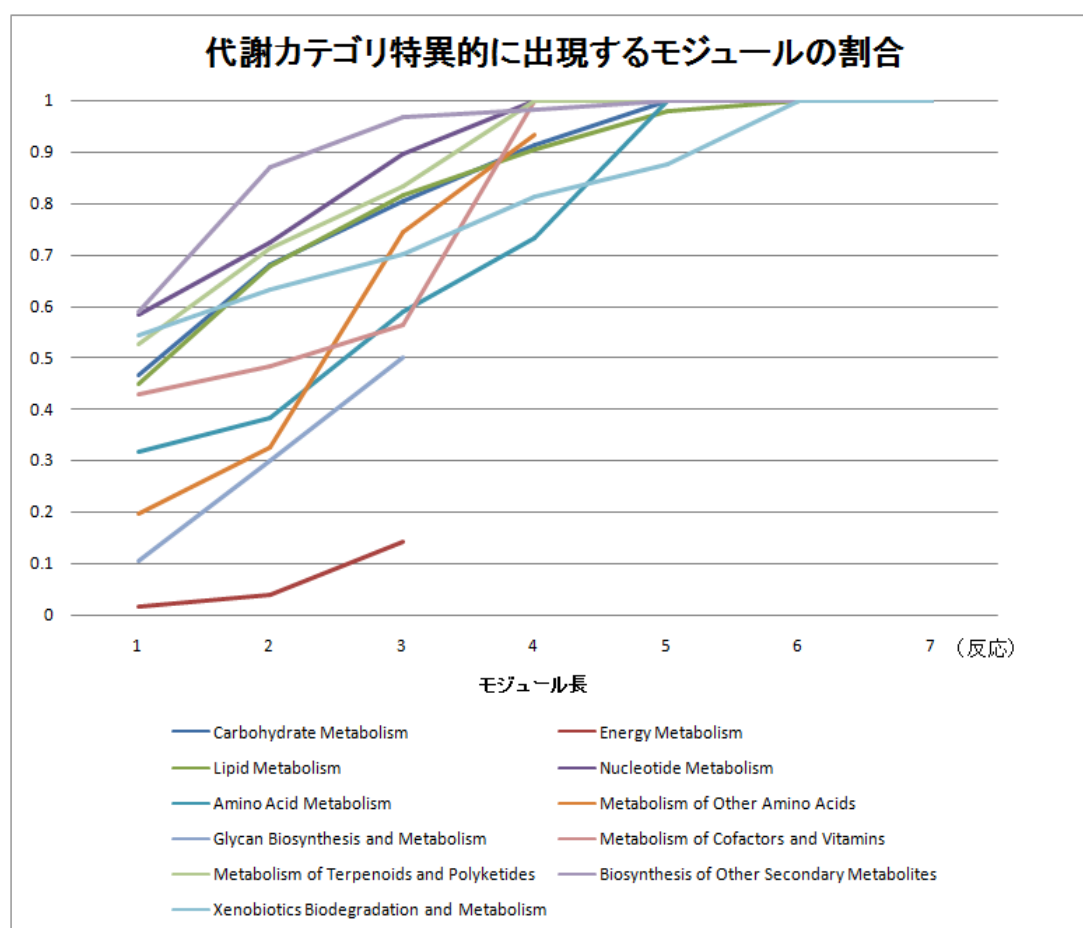


図 4-1 代謝カテゴリ特異的に出現する反応モジュールの割合

表 4-2 反応モジュールの一覧

それぞれの詳細は KEGG Reaction Module データベース

(<http://www.genome.jp/kegg/reaction/rmodule.html>)において公開されている。Reprinted with permission from [J Chem Inf Model.](#), (2013)53, 613-622. Copyright 2013 American Chemical Society.

Category	RC	Description	Length
Carboxylic acid metabolism			
2-Oxocarboxylic acid chain extension			
	RM001	2-Oxocarboxylic acid chain extension by tricarboxylic acid pathway	5
2-Oxocarboxylic acid chain modification			
	RM002	Carboxyl to amino conversion using protective N-acetyl group (basic amino acid syn	5
	RM032	Carboxyl to amino conversion	5
	RM033	Branched-chain addition (branched-chain amino acid synthesis)	3
	RM030	Glucosinolate synthesis	4
Fatty acid synthesis and degradation			
	RM018	Beta oxidation in acyl-CoA degradation	5
	RM019	Acyl-CoA synthesis and degradation via 3-hydroxy-acyl-CoA	4
	RM020	Fatty acid synthesis using acetyl-CoA (reversal of RM018)	4
	RM021	Fatty acid synthesis using malonyl-CoA	4
Aromatics degradation			
Methyl to carboxyl conversion on aromatic ring (preprocessing module)			
	RM003	Methyl to carboxyl conversion on aromatic ring	3
	RM015	Methyl to carboxyl conversion on aromatic ring, anaerobic	6
Dihydroxylation of aromatic ring			
	RM004	Dihydroxylation of aromatic ring, type 1 (dioxygenase and dehydrogenase reactions)	2
	RM005	Dihydroxylation of aromatic ring, type 1a (dioxygenase and decarboxylating dehydro	2
	RM006	Dihydroxylation of aromatic ring, type 2 (two monooxygenase reactions)	2
	RM007	Dihydroxylation of aromatic ring, type 3 (dealkylation and monooxygenase reactions)	2
Cleavage of aromatic ring			
	RM008	Ortho-cleavage of catechol (beta-ketoadipate pathway)	4
	RM009	Meta-cleavage of catechol	6
	RM013	Ortho-cleavage of halogenated aromatic ring	3
	RM016	Ring cleavage via beta oxidation, anaerobic	3
	RM017	Ring cleavage via Baeyer-Villiger oxidation	5
Dihydroxylation and cleavage of aromatic ring			
	RM010	Dihydroxylation and meta-cleavage of aromatic ring, type 1	4
	RM011	Dihydroxylation and meta-cleavage of aromatic ring, type 1b	3
	RM012	Dihydroxylation and meta-cleavage of aromatic ring, type 3a	5
	RM014	Ring removal from polycyclic aromatic ring	7
Other			
Amino acid metabolism			
	RM025	Conversion of amino acid moiety to carboxyl group (biogenic amine metabolism)	3
Nucleotide metabolism			
	RM024	Pyrimidine degradation	3
Sugar metabolism			
	RM022	Nucleotide sugar biosynthesis, type 1	3
	RM023	Nucleotide sugar biosynthesis, type 2	2
	RM034	Sugar degradation to aldehyde and pyruvate	4
Aromatics metabolism			
	RM027	Hydroxylation and methylation motif	2
	RM026	Hydroxylation and decarboxylation motif	2
	RM028	Flavonoid synthesis	3
	RM029	Pterocarpan synthesis	3
	RM031	Oxime to acetate conversion	3

4.3 反応モジュールの一般的特性

表 4-2 に示した反応モジュールについて 3 つの一般的特性が見出された。第一に、反応モジュールは異なる化学物質を生成する異なるパスウェイで繰り返し出現していた。第二に、反応モジュールは代謝ネットワークの基本構成要素であるかのように組み合わせられて用いられていることが分かった。第三に、もっとも重要なことには、反応モジュール (RC モジュールと呼ぶ) は基質・生成物の化学構造変化という化学的特性のみから得られたにも関わらず、ゲノム上の酵素のオーソログ遺伝子クラスターにより定義された KEGG パスウェイモジュール (KO モジュールと呼ぶ)、特に酵素群をコードするオペロン様構造の遺伝子クラスターと対応する傾向があった。言い換えれば、我々が導入部で言及した作業仮説「ゲノム上で連続して存在する遺伝子クラスターによって明らかにされた遺伝的機能単位は、有機反応の化学的機能単位を反映している」と矛盾しない結果が得られたと言える。次節で、表 4-2 のうち Carboxylic acid metabolism と Aromatics degradation について詳しく述べる。

4.4 各化学反応モジュールの生化学的意味付け

4.4.1 2-オキシカルボン酸鎖の伸長

最も特徴的な反応モジュールの一つが、重要な前駆体代謝化合物クラスである 2-オキシカルボン酸の主鎖伸長モジュール RM001 であった (図 4-2)。このモジュールは TCA 回路中のクエン酸やその他のトリカルボン酸に関係する良く

知られた連続反応列に対応しており、アセチル CoA 中のアセチル基の炭素がオキサロ酢酸(oxaloacetate)から 2-オキソグルタル酸へ、つまり 4 炭素化合物(C4)から 5 炭素化合物(C5)へと 2-オキシカルボン酸鎖を伸長するのに用いられる (図 4-2:i)。実際にこのモジュールはトリカルボン酸を含む TCA 回路の一部として抽出された。興味深いことに、我々は同じ RM001 反応モジュールを異なる 3 つのパスウェイから同定した。一つはリジン生合成パスウェイ(map00300)における 2-オキソグルタル酸(C5)の 2-オキソアジピン酸(C6)への更なる伸長である (図 4-2:ii)。もう一つはバリン、ロイシン、イソロイシン生合成経路(map00290)で見つかる、ピルビン酸の 2-オキソブタン酸への伸長 (図 4-2:iii)、及び 2-オキソイソ吉草酸の 2-オキソカプロン酸への伸長である (図 4-2:iv) [25]。更に、植物二次代謝化合物であるグルコシノレートの合成パスウェイ(map00966)では、2-オキソ-4-メチルチオブタン酸から 2-オキソ-10-メチルチオデカン酸までの経路に RM001 モジュールが 6 回、繰り返し並んでいることが見つかった。クエン酸回路の KEGG パスウェイマップでのオキサロ酢酸から 2-オキソグルタル酸への変換(RM001)は次のように表される：まずオキサロ酢酸とアセチル CoA がクエン酸を生成し(RC00067)、次に cis-aconitate へと変換され(RC00498)、さらにイソクエン酸へと変換され(RC00618)、最後に 2 反応ステップ(RC00084+RC00626)もしくはその全体反応(RC00114)で 2-オキソグルタル酸へと変換される (図 4-2:i)。このように、最後の変換が 2 ステップの連続反応として表されたり全体反応として表されたりするのは、EC 番号が必ずしも反応だけに着目した分類法ではないことによる。同じ構造変換を起こす反応であっても、EC 1.1.1.42 (isocitrate

dehydrogenase (NADP+))では 2 ステップの反応が詳細に記述され、EC 1.1.1.41 (isocitrate dehydrogenase (NAD+)) はそれらを 1 つの反応とみなした反応式しか定義されていない(RC00114)(図 4-3)。同様に、2 番目と 3 番目の反応(RC00498+RC00618)は同一の酵素 EC 4.2.1.3 によって触媒されるひとつの反応(クエン酸からイソクエン酸への 1 ステップの反応)と見なすこともできるが、このように詳細ステップが明らかになっている反応は、全体反応よりも各反応ステップに対し R 番号や RC 番号などの ID が割り当てられているため、KEGG の代謝パスウェイ中には含まれていない。表 4-3 に示すように、RM001 反応モジュールは明らかに異なる RCLASS 配列から成るが、これらの背景を考慮すると実際には同じものと言うことが出来る。

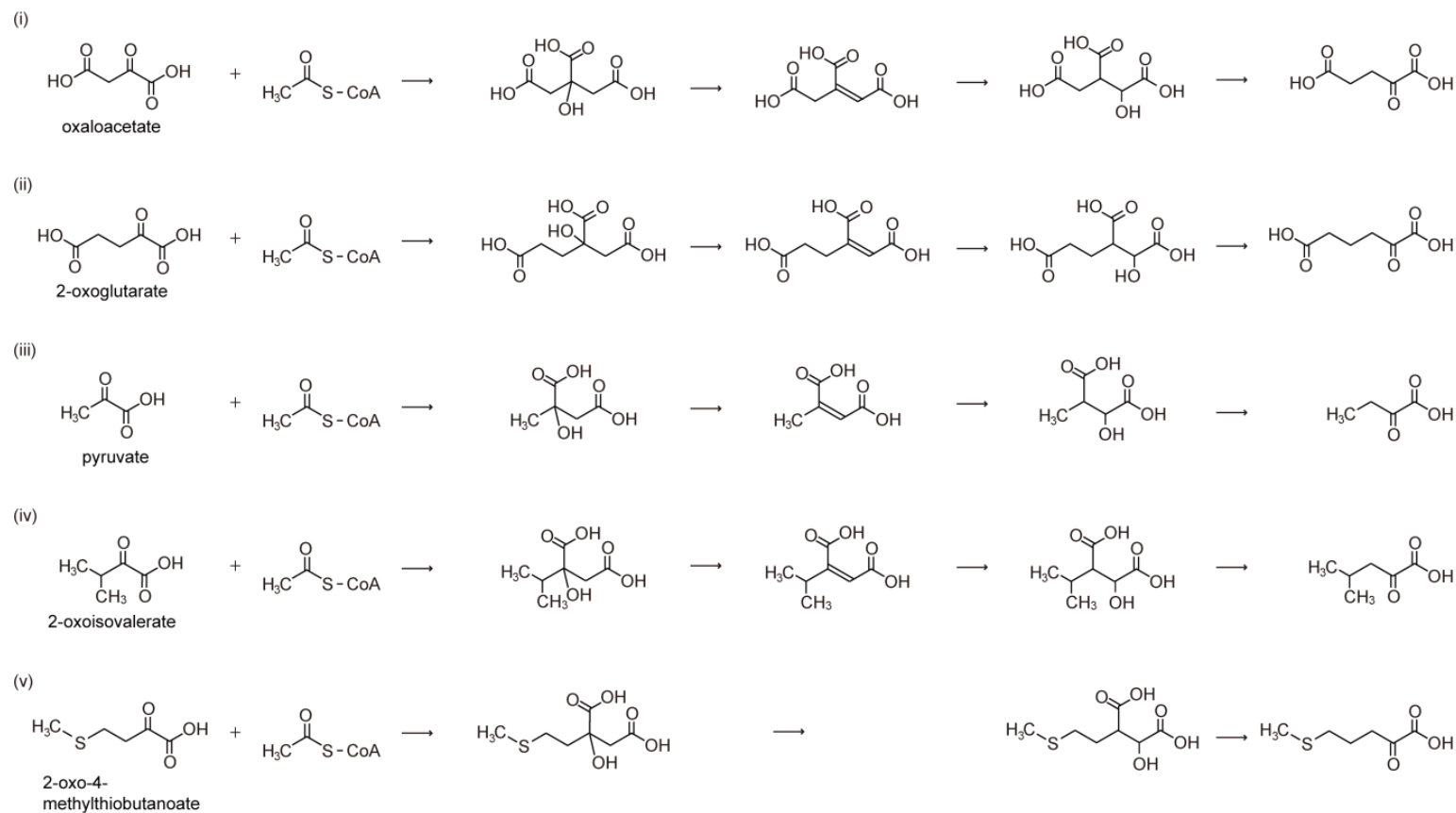
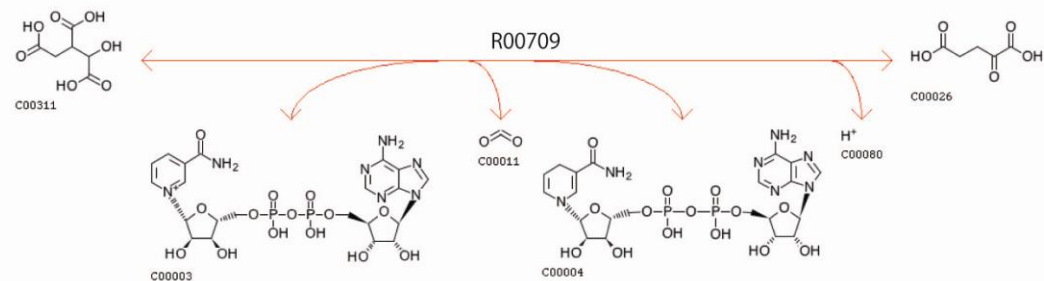


図 4-2 2-オキソカルボン酸伸長モジュール (RM001)

2-オキソカルボン酸伸長モジュールは 1 分子の acetyl-CoA を消費して 2-オキソカルボン酸鎖を伸長する。(i)オキサロ酢酸から 2-オキソグルタル酸への伸長 (ii)2-オキソグルタル酸から 2-オキソアジピン酸への伸長 (iii)ピルビン酸の 2-オキソブタン酸への伸長 (iv)2-オキソイソ吉草酸の 2-オキソカプロン酸への伸長. (v) 2-オキソ-4-メチルチオブタン酸から 2-オキソ-5-メチルチオデカン酸までの伸長 Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright 2013 American Chemical Society.

EC 1.1.1.41



EC 1.1.1.42

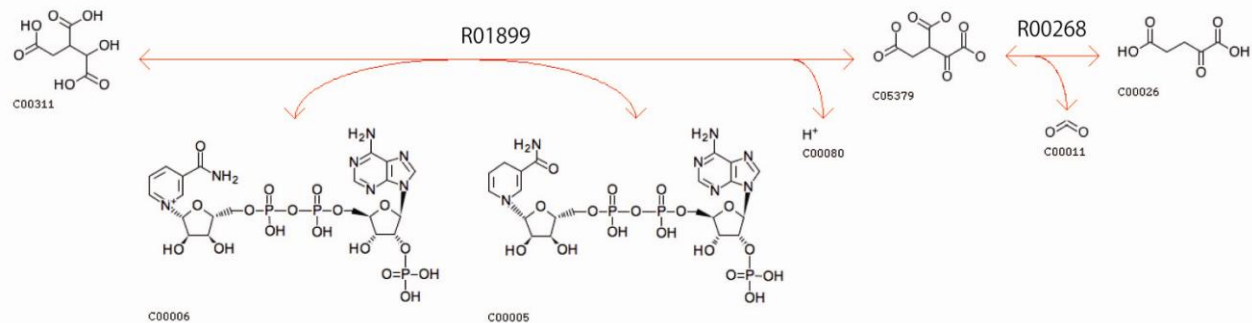


図 4-3 EC 1.1.1.42 と EC 1.1.1.41

EC 1.1.1.42 (isocitrate dehydrogenase (NADP+))では 2 ステップの反応が詳細に記述され、EC 1.1.1.41 (isocitrate dehydrogenase (NAD+)) はそれらを 1 つの反応とみなした反応式しか定義されていない(RC00114) Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622.](#)
Copyright 2013 American Chemical Society.

表 4-3 2-オキソカルボン酸関連反応モジュール群

Reprinted with permission from [J Chem Inf Model.](#), (2013)53, 613-622. Copyright 2013 American Chemical Society.

RC module	Pathway	Overall reaction	RCLASS sequence
RM001	Citrate cycle (map00020)	oxaloacetate → 2-oxoglutarate	RC00067 RC00498 RC00618 RC00084+RC00626
	Lysine biosynthesis (map00300)	2-oxoglutarate → 2-oxoadipate	RC00067 RC00498 RC00618 RC00114
	Isoleucine biosynthesis (map00290)	pyruvate → 2-oxobutanoate	RC01205 RC00976 RC00977 RC00417
	Leucine biosynthesis (map00290)	2-oxoisovalerate → 2-oxoisocaproate	RC00470 RC01041 RC01046 RC00084+RC00577
	Glucosinolate biosynthesis (map00966)	2-oxo-4-methylthiobutanoate → 2-oxo-10-methylthiodecanoate	RC00067 RC00497 RC00114 (six repeats)
RM002	Lysine biosynthesis (map00300)	2-aminoadipate → lysine	RC00064 RC00043 RC00684 RC00062 RC00064
	Arginine biosynthesis (map00330)	glutamate → ornithine	RC00064 RC00043 RC00684 RC00062 RC00064
RM032	Ectoine biosynthesis (map00260)	aspartate → 2,4-diaminobutanoate	RC00043 RC00684 RC00062
RM033	Valine biosynthesis (map00290)	pyruvate → 2-oxoisovalerate	RC01192 RC00837 RC00726 RC00468
	Isoleucine biosynthesis (map00290)	2-oxobutanoate → 3-methyl-2-oxopentanoate	RC01192 RC01726 RC00726 RC01714
RM030	Glucosinolate biosynthesis (map00966)	homomethionine → glucoiberberin	RC02295 RC02210 RC02265 RC00882 RC00883

4.4.2 2-オキシカルボン酸鎖の修飾

2-オキシカルボン酸の代謝系について反応モジュールを詳しく調べると、トリカルボン酸パスウェイによる 2-オキシカルボン酸鎖の伸長モジュール RM001 は、還元的アミノ化ステップ(RC00006 又は RC00036)と、それに連続して起こる RM002 (類似反応モジュールである RM032 を含む)、RM033 及び RM030 の 3 つの修飾モジュールとの組み合わせで用いられていることがわかった。図 4-4 に示す、種々のアミノ酸を生成する反応モジュールの構造はその一例である。

カルボキシル基をアミノ基に変換する反応モジュール RM002 (類似反応モジュールである RM032 を含む) はオルニチン関連アミノ酸生合成において、分岐鎖を付加するための反応モジュール RM033 は分岐アミノ酸(バリン、ロイシン、イソロイシン)生合成において、ともに RM001 との組み合わせで様々なアミノ酸を生成する。グルコシノレート生合成パスウェイにおいては、オキシムへの変換ののちチオ-グルコース部分構造を付加する反応モジュール RM030 が RM001 と組み合わせあって様々なグルコシノレートを生成する。(図 4-5)

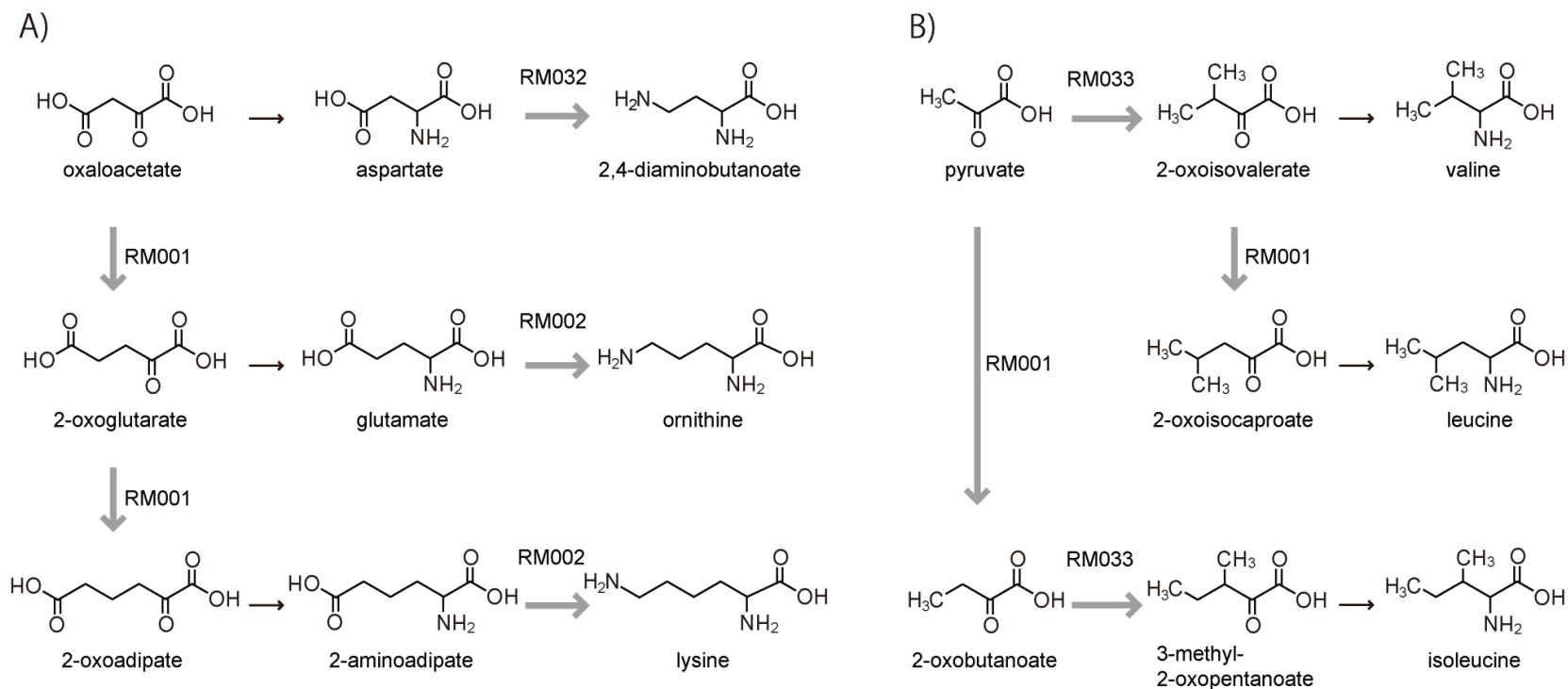


図 4-4 アミノ酸生成モジュール

A)オルニチン関連アミノ酸、B)分岐アミノ酸を生成する、2-オキソカルボン酸鎖伸長及び修飾からなる反応モジュール構造。垂直方向の矢印は伸長モジュールRM001を示す。水平方向の矢印は還元アミノ化ステップ(RC00006又はRC00036)と連続する修飾モジュールRM002, RM032及びRM033を示す。Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright 2013 American Chemical Society.

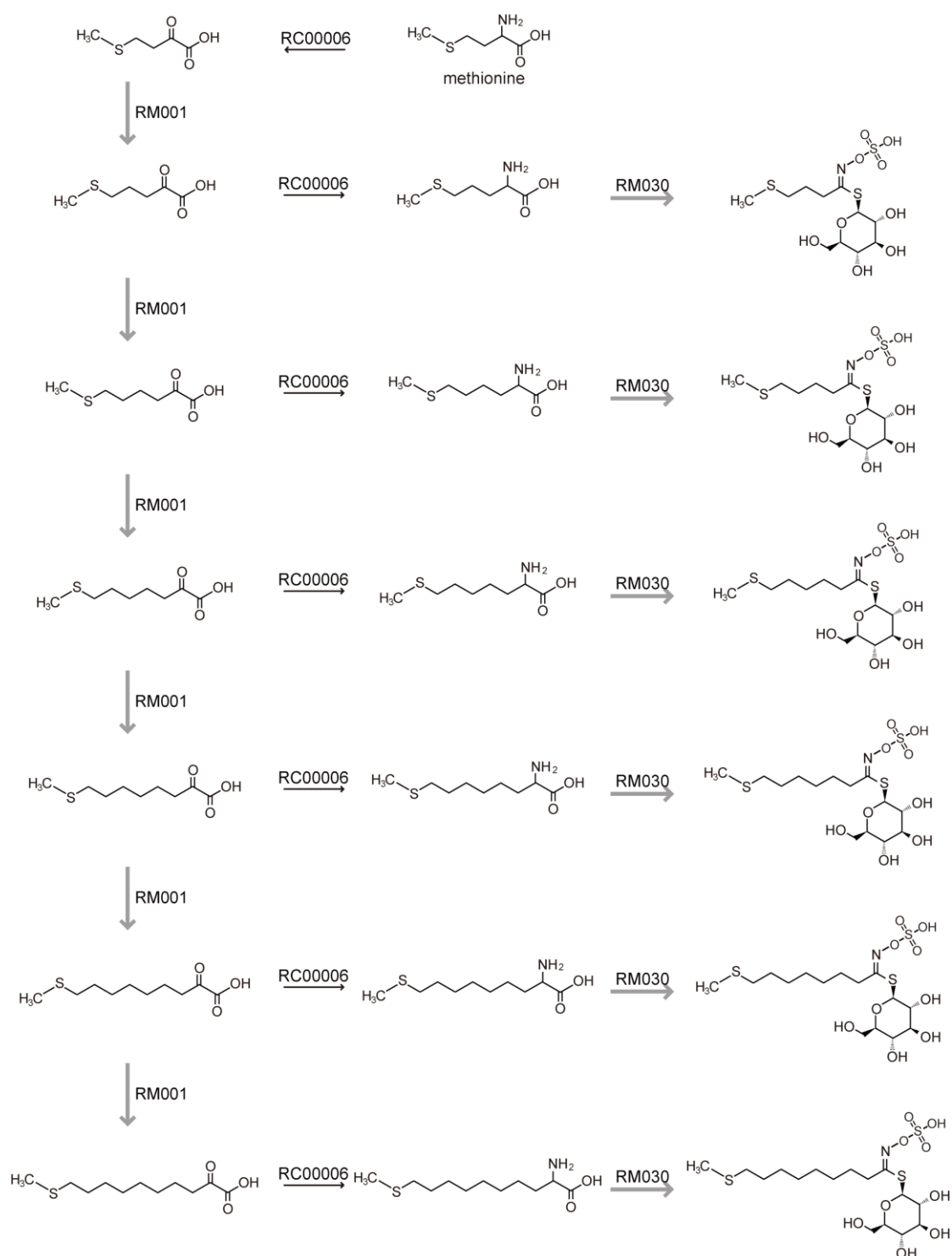


図 4-5 グルコシノレート生合成パスウェイモジュール

RM001 (オキシカルボン酸関連反応モジュール) と **RM030** (オキシムへの変換ののちチオ-グルコース部分構造を付加する反応モジュール) を含むメッシュ様構造。**RC00006** と **RC00036** は還元アミノ化に対応する。Reprinted with permission from [J Chem Inf Model.](#) (2013)53, 613-622. Copyright 2013 American Chemical Society.

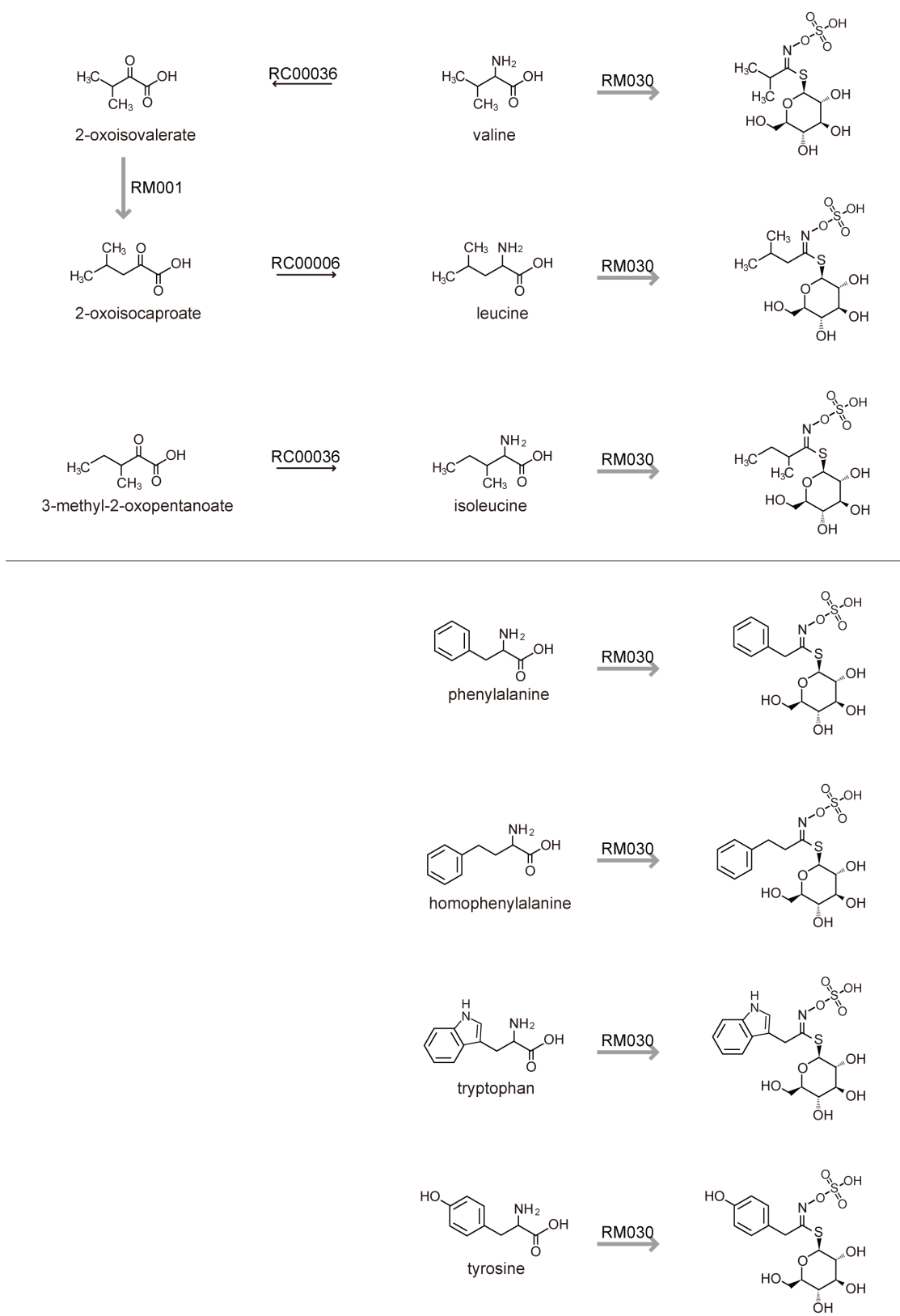
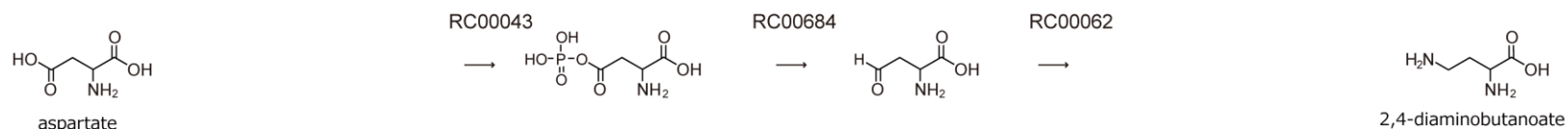


図 4-5 (続き)

図 4-6 に表 4-3 と図 4-4 中で述べた化学修飾モジュール RM002 と、類似反応モジュールである RM032 の詳細を示す。RM002 を構成する連続反応配列の最初と最後のステップ(RC00064)は保護基としての N-アセチル基の修飾及び脱修飾反応であった。我々は同じ変換反応を行うが保護基の修飾・脱修飾を含まない反応モジュール(RM032)を、エクトイン生合成パスウェイ(map00260) 中の、アスパラギン酸塩を 4-ジアミノブタン酸へ変換する経路 (図 4-6) に発見した。これらの経路は反応モジュールの多様性あるいは進化を示唆するものであり興味深い。RM032(RC00043 RC00684 RC00062) が短鎖カルボン酸鎖に使われるのに対し、RM032 の前後に N アセチル化と N 脱アセチル化ステップを持つ RM002 (RC00064 RC00043 RC00684 RC00062 RC00064)はより長いカルボン酸鎖に対して使われていた。さらに、2-アミノアジピン酸をリジンへ変換するリジン合成パスウェイでは、保存反応列を囲むステップは (アセチル基による修飾ではなく) 酵素タンパク質の N 末端による修飾になっていた[25]。つまり、全体を眺めると同一に見える連続反応であっても、その連続反応は基質の種類に応じて、修飾基が不要であったり、簡単な修飾基 (アセチル基) を用いていたたり、複雑な修飾基 (キャリアタンパク質) を用いていたたりするなど、異なる複雑さによって達成されていることが明らかになった。これは長い鎖を持つ基質に対処するための代謝パスウェイの化学進化を示しているのかもしれない。

RM032



RM002

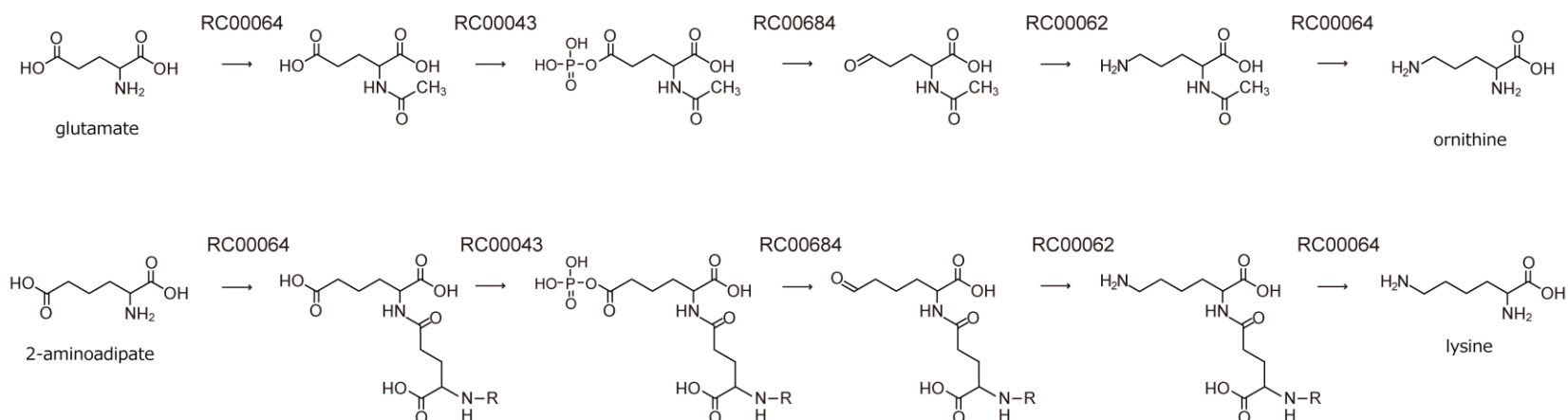


図 4-6 カルボキシル基をアミノ基へ変換するモジュール RM002 と RM032

短い主鎖の化合物を基質とするモジュール RM032 と、より長鎖の化合物を基質とするモジュール RM002 の中央の 3 反応は同一の反応列である。RM002 はタンパク質の N 末端の N-アセチル基の付加反応と脱離反応(RC00064)を両端に持つことで、RM002 よりも 2 反応長いモジュールとなっている。このアセチル基は基質の自発的な分子内環化を妨げるための、保護基としてはたらいっていると考えられる。

Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright 2013 American Chemical Society.

4.4.3 脂肪酸合成と β 酸化

脂肪酸は長鎖炭化水素の 1 価のカルボン酸である。脂肪酸の合成はケトアシル合成酵素(KS)、ケトリダクターゼ(KR)、デヒドラーゼ(DH)及びエノイルリダクターゼ(ER)反応の 4 段階の連続反応列の繰り返しであり、4 反応ごとに炭化水素鎖を炭素 2 つ分ずつ伸長する。生物種によってこの 4 つの酵素の遺伝子構成は異なり、4 つの独立した遺伝子を持つ生物、KS+KR ドメインと DH+ER ドメインから成る 2 つの遺伝子を持つ生物、もしくは全てのドメインからなる一つの遺伝子を持つ生物が知られている。よって、脂肪酸合成遺伝子の構成には遺伝学的な面からみた進化的関連性が考えられる。脂肪酸合成には主経路と副経路があり、アシルキャリアータンパク質が関与するマロニル CoA を炭素源とする主経路(map00061)に対し、ミトコンドリアで起こる副経路(map00062)ではアシルキャリアータンパク質が関与せず、アセチル CoA を炭素源として使用する。この副経路は後述する脂肪酸分解で現れる β 酸化のモジュール(RM018)の逆反応と見なすことが出来る。これら主経路及び副経路にあたる反応モジュール RM021 及び RM020 は、以下の RCLASS 配列として得られた。(図 4-7)

RM021: (RC02729 RC00117 RC00831 RC00076) for $n = 2$
(RC02728 RC00029 RC01095 RC00052) for $n \geq 4$

RM020: (RC00326 RC00117 RC00831 RC00076) for $n = 2$
(RC00326 RC00029 RC01095 RC00052) for $n \geq 4$

ここでマロニル CoA とアセチル CoA を区別する KS 反応ステップ(RC02729, RC02728 及び RC00326)以外は、RM021 と RM020 は完全に同一であった。我々はアシルキャリアータンパク質の関与は、上述の「2-オキソカルボン酸の修飾」と同様、反応の特異性と効率性を高める脂肪酸合成の化学的観点からみた進化であると考えている。

β 酸化モジュール RM018 (RM020 の逆反応) は、次の RCLASS 配列から成っていた(図 4-8) :

RM018: (RC00052 RC01095 RC00029 RC00326) for $n \geq 4$
(RC00076 RC00831 RC00117 RC00326) for $n = 2$

脂肪酸分解経路(map00071)だけでなく、RM018 はカプロラクタム分解(map00930)におけるアジピル CoA からスクシニル CoA への反応や、イソロイシン分解(map00280)における 2-メチルブタノイル CoA からプロパノイル CoA への反応にも見つかった。様々な開裂反応に関わる RM018 類似反応列は、嫌気性安息香酸分解パスウェイ(map00362)におけるピメロイル CoA からグルタリル CoA への反応や、一次胆汁酸合成(map00120)においてケノデオキシコロイル CoA やコロイル CoA への反応にもみられた。

より興味深いことに、類似性のやや低いモジュール RM016 (RC02034 RC00154 RC01429)(図 4-8)は嫌気性安息香酸分解パスウェイ(map00362)及びリモネン・ピネン分解パスウェイ(map00903)にも見つけることができた。RM016 はアシル鎖切断 (チオール開裂) というよりはむしろ芳香族環開裂(加水分解)のモジュールであるが、開裂反応に先だって酸素を付加する戦略は類似していた。RC02034 と RC01095 (又は RC00831)の類似性スコアは 0.8 であり、RC00154 と RC00029 (又は RC00117)の類似性スコアは 0.7 であった。さらに、嫌氣的安息香酸分解パスウェイにおける RC00154 の遺伝子と脂肪酸合成パスウェイにおける RC00029 では遺伝子間にも類似性が見られた[26]。

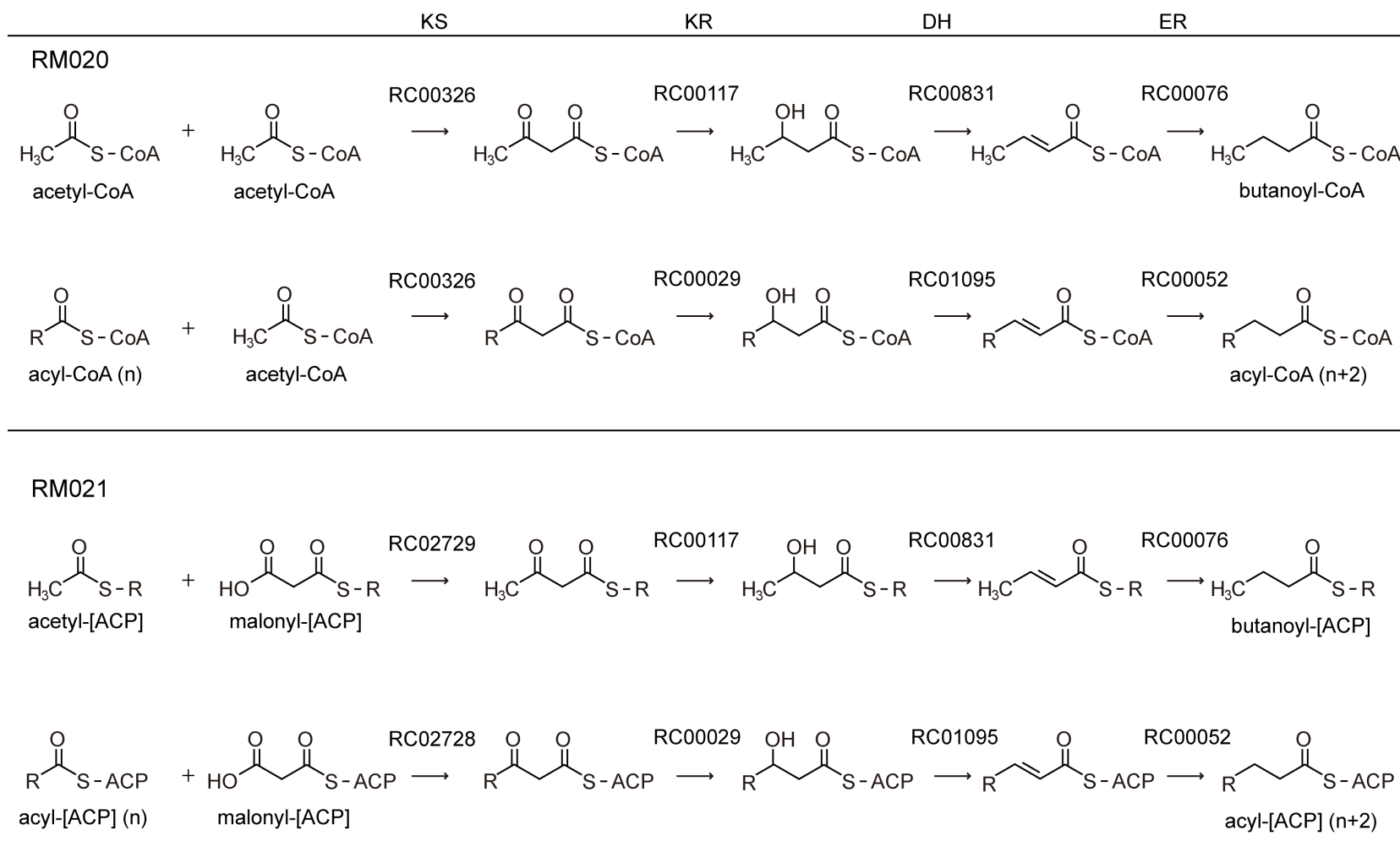


図 4-7 脂肪酸生合成モジュール RM021 (主経路) と RM020 (副経路)

Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright 2013 American Chemical Society.

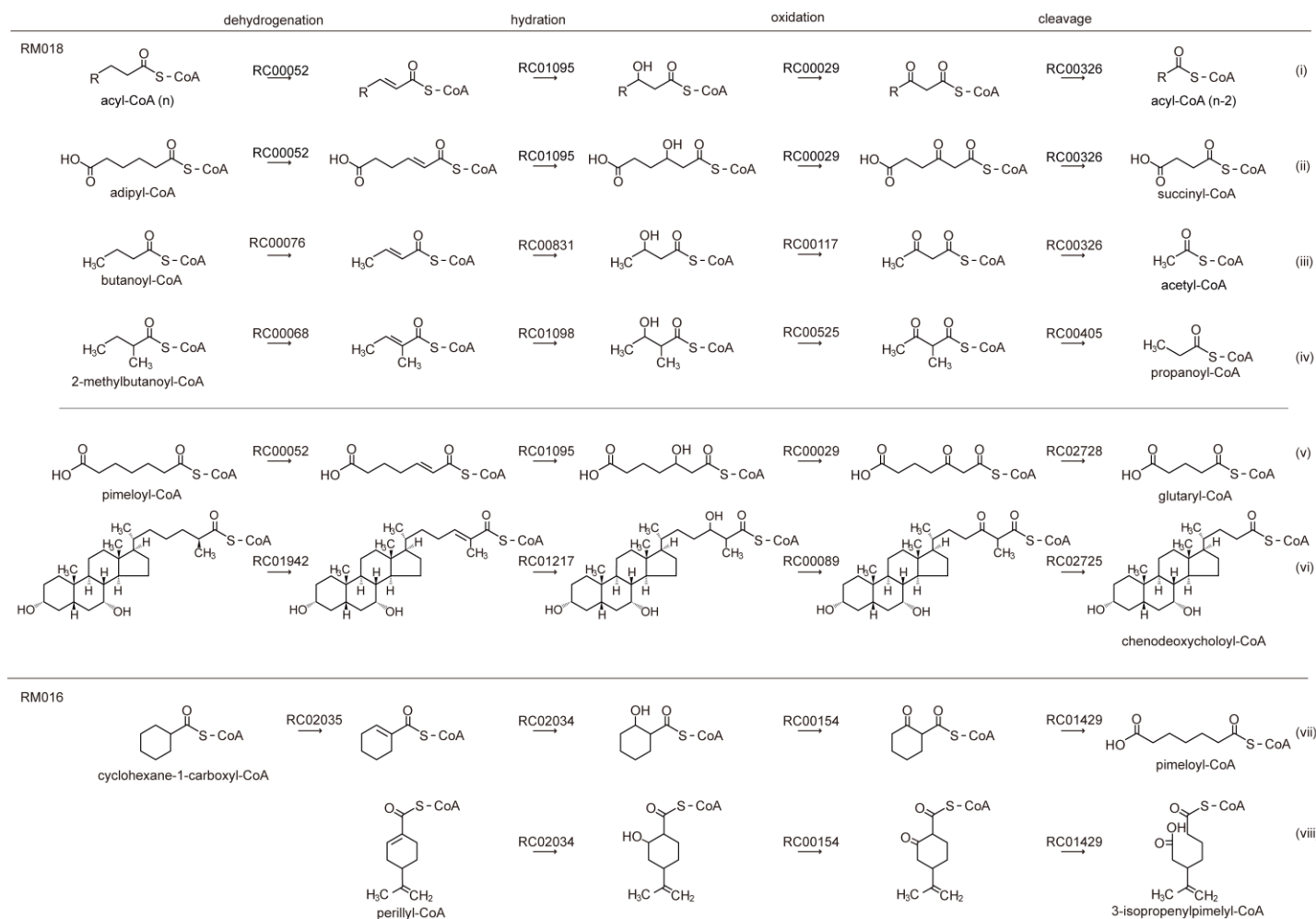


図 4-8 β酸化モジュール

β酸化モジュール RM018 は脱水素化、水和反応、酸化、開裂反応の 4 反応からなる。(i, iii) 脂肪酸分解パスウェイ (map00071) に見られる RM018 モジュール (ii) カプロラクタム分解パスウェイ (map00930), (iv) イソロイシン分解パスウェイ (map00280), (v) 安息香酸エステル分解パスウェイ (map00362), and (vi) 一次胆汁酸生合成パスウェイ (map00120)。RM016 は良く似たモジュールで、環状化合物に対して働く。(vii) 安息香酸エステル分解パスウェイ (map00362) と (viii) リモネン・ピネン分解パスウェイ (map00903) から得られている。Reprinted with permission from [J Chem Inf Model.](#), (2013)53, 613-622. Copyright 2013 American Chemical Society.

4.4.4 微生物の生分解パスウェイにおける芳香族環分解

微生物は環境中の人工化合物を含む幅広い化学物質の分解能を持つことが知られている。本研究の解析の結果、環ジヒドロキシル化モジュール(RM004-RM006)や環開裂モジュール(RM008 及び RM009)、そして前処理モジュール(RM003)の芳香族環分解モジュールが得られた (図 4-9)。前処理モジュールは例えば芳香族環に隣接したメチル基をカルボキシル基に変換するなど、置換基の酸化に働く。RM003 はモノオキシゲナーゼ反応(EC 1.14.13)とデヒドロゲナーゼ反応(EC 1.1.1 及び 1.2.1)の 2 段階反応からなり、図中のトルエンから安息香酸、p-シメンから p-クミン酸への変換に示すように、それに環ジヒドロキシル化モジュールが続く。

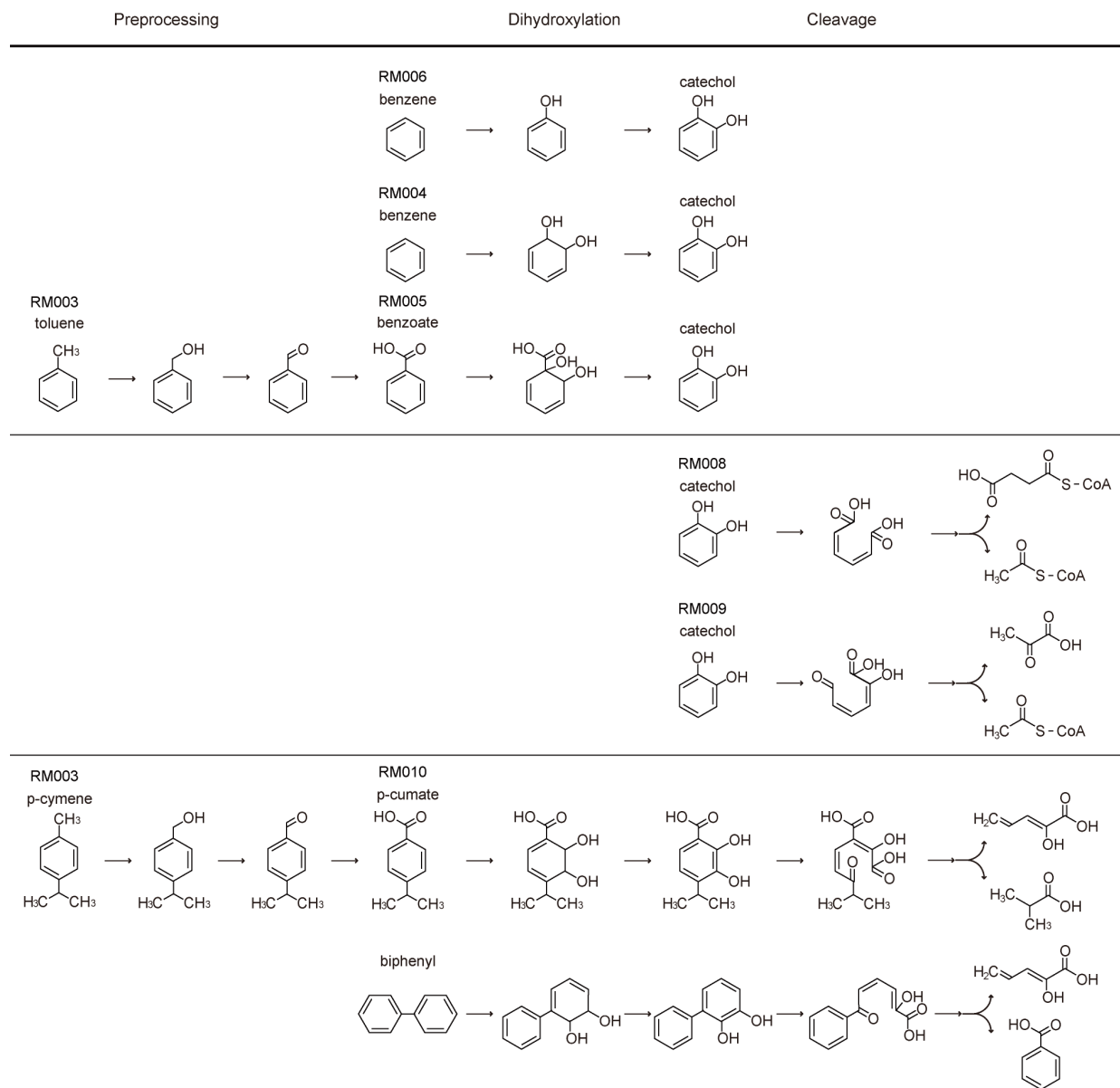


図 4-9 微生物分解パスウェイを構成する芳香族環分解モジュール

芳香族環は次の 3 つの段階（反応モジュール）を経て分解される。最初のステップは補助的な前処理段階(RM003) であり、芳香族環上のメチル基をカルボキシル基に変換する。第二段階はジヒドロキシル化のステップであり、芳香族環に二つのヒドロキシル基がどのように付加されるかにより 3 つのタイプ(RM006, RM004 及び RM005)に分類される。第三段階は芳香環が開裂する場所によってオルト開裂(RM008)とメタ開裂(RM009)のどちらかに分類され、それぞれに TCA 回路代謝中間体へと至る。

Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright 2013 American Chemical Society.

RCLASS エントリの類似性グループに基づき、我々はジヒドロキシ化反応のステップを 3 つのモジュールにカテゴリ分けした(RM004、RM005 及び RM006: 図 4-9)。RM004 (type 1)と RM005 (type 1a) がジオキシゲナーゼ反応に続くデヒドロゲナーゼ反応を用いていたのに対し、RM006 (type 2)は 2 ステップのモノオキシゲナーゼ反応を利用していた。前処理モジュール RM003 の後続く例の多い RM005 モジュールでは、ヒドロキシ基の一つはカルボキシ基の α 位に付加され、続くデヒドロゲナーゼ反応時に脱カルボキシ化も起こる。それぞれのジヒドロキシ化反応モジュールの後には、二つのヒドロキシ基のメタ位又はオルト位のどちらかで起こるヒドロキシ環開裂ジオキシゲナーゼ反応が続く。

RM008 及び RM009 はそれぞれ、カテコールや関連化合物のオルト及びメタ開裂モジュールに相当し、開裂反応とそれに続く TCA 回路中間体を生成する反応を含んでいた。オルト開裂パスウェイは β ケトアジピン酸パスウェイ[27]として知られており、カテコールはスクシニル CoA とアセチル CoA に変換される。それに対しメタ開裂パスウェイではカテコールはピルビン酸とアセチル CoA に変換される。置換基の大きな分子の場合オルト開裂が立体障害等の影響で難しいこともあるのに対し、メタ開裂はより大きな分子にも反応しやすいと考えられ、より広い基質に対応できる一般的な分解戦略であると考えられる。我々はジヒドロキシ化モジュール RM004 とカテコールメタ開裂モジュール RM009 の一部から構成される大きなモジュールを RM010 反応モジュール(図 4-9)と定義した。trans-ケイ皮酸、p-クミン酸、エチルベンゼン、スチレン及びダイオキシンを含む様々な化合物が、このモジュールによって RM009 によるさらなる処理を受ける

2-オキソペント 4- エン酸へと分解される。例えば、二つの環をもつビフェニルは安息香酸へと部分処理され、その後他の分解モジュールによってさらなる分解を受ける。

これらの環のヒドロキシ化によるカテコール形成を経た芳香族環分解モジュールは好氣的経路において起こると考えられるが、我々は嫌氣的経路からも芳香族環分解モジュールを検出した。前述の β 酸化様モジュール RM016 は芳香族環分解モジュールの嫌氣性版[28]である。我々はメチル酸化前処理モジュールの CoA を利用する嫌氣性版 RM015 も同定した（トルエン分解における好氣性モジュール RM003 と嫌氣性モジュール RM015 の比較は図 4-10 に示した）。

RM003 Aerobic



RM015 Anaerobic

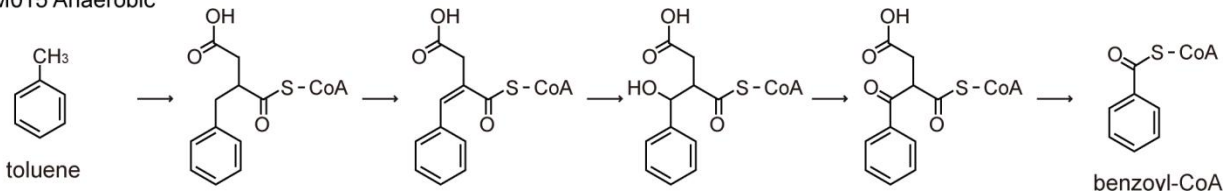


図 4-10 トルエン分解における好氣性モジュール RM003 と嫌氣性モジュール RM015 の比較

Reprinted with permission from [J Chem Inf Model.](#), (2013)53, 613-622. Copyright 2013 American Chemical Society.

表 4-4 KEGG モジュールに対応する生分解反応モジュール

Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright 2013 American Chemical Society.

RC module	KO module	Overall reaction
RM003	M00538	toluene → benzoate
	M00537	o-xylene → o-methylbenzoate
	M00419	p-cymene → p-cumate
RM004	M00547	benzoate → catechol
RM005	M00551	benzoate → catechol
RM006	M00548	benzene → catechol
RM008	M00568	catechol → 3-oxoadipate
RM009	M00569	catechol → pyruvate + acetaldehyde
RM010	M00539	p-cumate → 2-oxopent-4-enoate + methylpropanoate
	M00543	biphenyl → 2-oxopent-4-enoate + benzoate
RM011	M00544	carbazole → 2-oxopent-4-enoate + anthranilate
RM015	M00418	toluene → benzoyl-CoA

ベンゼン、トルエン、エチルベンゼンやキシレン(BTEX 化合物)などの生体異物の生分解能は主にプラスミド上に適切な遺伝子セットを持つ特定の生物種に限られている。よって、ここで同定された反応モジュールはゲノム上のオペロン様遺伝子クラスターとの対応が期待された。表 4-4 に、KEGG モジュール (KO modules) 多数がオペロン様遺伝子クラスターにコードされているときの、微生物の生分解パスウェイの反応モジュール (RC modules) と KEGG モジュールの対応を示す(各々の KEGG モジュールに属する遺伝子は、<http://www.kegg.jp/module/M00548> などそれぞれのエントリのウェブページにある「Ortholog table」というリンクから閲覧できる)。例えば、*Pseudoxanthomonas spadix* BD-a59[29] (KEGG organism code: psd, T01643)は、トルエン及びキシレンの各前処理モジュール M00538 及び M00537、ジオキシゲナーゼ及びモノオキシゲナーゼ触媒ジヒドロ

キシル化モジュール M00548、メタ開裂モジュール M00569、そしてダイオキシン分解モジュール M00543 及び M00544 を持つため、BTEX 分解能を持つことを示している。

4.5 酵素遺伝子クラスターにコードされる反応モジュール

KEGG MODULE データベース中の KEGG パスウェイモジュール(KO モジュール)は、KEGG データベースを維持管理する専門キュレーターにより手作業で整理された、酵素オーソログ遺伝子のセットであり、原核生物ゲノム中でオペロン様遺伝子クラスターとの対応なども考慮されている。我々は本解析で抽出された反応モジュール(RC モジュール)と、既に定義されていた KO モジュールとの関係を調べた。KO モジュールについては手作業による定義の揺れを除くため、遺伝子クラスター（ゲノム上の並びが保存されているもの）を構成するものだけに限定した。その結果、例えば反応モジュール RM001 は KO モジュール M00010、M00432 及び M00535 とよく対応する生物種が多いことを発見した。表 4-5 に示すように、*Pyrococcus furiosus*[30]のゲノムでは二つの遺伝子クラスターが反応モジュール RM001 に対応する：遺伝子クラスター(PF0203 PF0201 PF0202)はクエン酸回路の RCLASS 配列(RC00067 RC00498+RC00618 RC00084+RC00626)に対応し、遺伝子クラスター(PF0937 PF0938+PF0939 PF0940)はロイシン生合成の RCLASS 配列(RC00470 RC01041+RC01046 RC00084+RC00577)に対応する。最初の酵素（クエン酸合成酵素 PF0203 とイソプロピルリンゴ酸合成酵素 PF0937）については配列類似性が見られないにも拘らず、2 番目のヒドラターゼ(PF0201 及び PF0938+PF0939)と 3 番目のデヒドロゲナーゼ群(PF0202 及び PF0940)は 35%程度で配列が一致しておりパラログ遺伝子グループを形成していることは、遺伝的多様性と化学的多様性の繋がりを示唆すると考え

られ興味深い(図 4-11)。表 4-5 には RM001 に加えて RM002 の酵素遺伝子クラスターの例のごく一部を示した。他の多くの例は KEGG データベース中の KEGG MODULE エントリのオーソログテーブルで閲覧可能である。全てのエントリは <http://www.kegg.jp/module/M00010> 等からアクセス可能である。

表 4-5 酵素遺伝子クラスターと対応する反応モジュール

RC module	Overall reaction	KO module	Gene cluster example*
RM001	oxaloacetate → 2-oxoglutarate	M00010	(pfu)PF0203 PF0201 PF0202
	2-oxoisovalerate → 2-oxoisocaproate	M00432	(pfu)PF0937 PF0938+PF0939 PF0940
	pyruvate → 2-oxobutanoate	M00535	(bth)BT_1858 BT_1860+BT_1859 BT_1857
RM002	glutamate → ornithine	M00028	(bsu)BSU11200 BSU11210+BSU11190 BSU11220
	2-aminoadipate → lysine	M00031	(ttr)Tter_0315+Tter_0316 Tter_0320 Tter_0319 Tter_0321 Tter_0317

* KEGG データベース中で生物種はアルファベット 3 文字 (またはアルファベットの T とそれに続く 5 桁の数字) で次のように表されている : pfu (T00075), *Pyrococcus furiosus* DSM 3638; bth (T00122), *Bacteroides thetaiotaomicron* VPI-5482; bsu (T00010), *Bacillus subtilis* 168; ttr (T01134), *Thermobaculum terrenum* ATCC BAA-798. Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622.](#)
Copyright 2013 American Chemical Society.

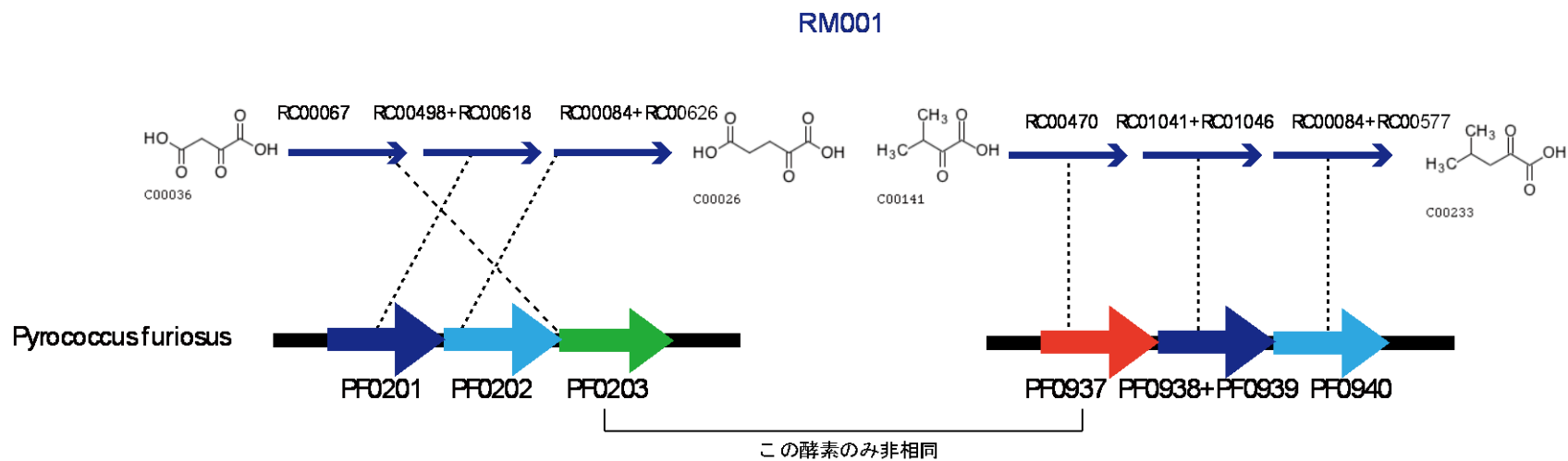


図 4-11 反応モジュール RM001 に対応する *Pyrococcus furiosus* の二つの遺伝子クラスター

第5章 考察

本解析では、代謝ネットワークが保存された反応列からなる化学的反応モジュール (RC modules) を含むこと、それらが様々な代謝系で組み合わせられて用いられていること、また伝統的な機能モジュール (KO modules) や原核生物ゲノム上の遺伝子クラスターとよく一致することを示した。代謝ネットワークのモジュール性についてはこれまでに様々な研究で言及されてきた[17, 31-32]が、我々の知る限り、本研究は代謝のネットワークについて化学的ネットワークと遺伝子ネットワークという二つの視点からみたモジュール性を報告し、それらの共進化を示唆した初めての例である。結果の項目で言及しなかったいくつかのモジュールの特徴についてここに簡単に述べる。

5.1 逆反応の利用

ネットワークは、エッジに向きの情報の与えられた有向グラフによるネットワークと、向きの情報を持たず繋がりだけを示す無向グラフによるネットワークに分けられる。代謝ネットワークは、可逆反応は両向きの、不可逆反応は片方向きの情報が与えられた有向グラフで表される。我々は、連続反応の保存性によって代謝ネットワークから化学的反応モジュールを抽出した。このとき連続反応を正しい向きでアラインメントするためには、反応だけでなくそれらを繋ぐエッジの向きについてもマッチングを行わなくてはならない。しかしながら、KEGG PATHWAY 上の反応は、約 2 割が可逆反応、約 8 割が不可逆反応として記載されている。代謝の流れから恐らく一方向に進むと考えられる経路であっても、可逆反応についてはデータベースから方向を読み取れないため、やむな

く可逆反応については両方向でマッチングを行った。しかしこのことにより、われわれは興味深い特徴を捕えることができた。高頻度で出現する反応モジュールに関しては、多くの場合その逆反応の反応モジュールが存在したのである。

例としてアミノ酸から生体アミンやその他の重要な代謝化合物を生成する反応モジュール **RM025** を挙げる。(図 5-1)リン酸化のステップ(**RC00043**)に続きカルボキシル基をアミノ基に変換する(**RC00684 RC00062**)モジュール **RM002**(保護基あり)と **RM032**(保護基なし)については既に 4 章で述べた。(図 4-6)**RM025** の脱カルボキシル基ステップ(**RC00299**)に続くアミノ基をカルボキシル基に変換する 2 反応(**RC00062 RC00080**)は、**RM032** のカルボキシル基をアミノ基に変換する 2 反応(**RC00684 RC00062**)の逆反応配列である。**R00080** はフィンガープリント表現では **RC00684** と同一であり、アミノ基をカルボキシル基に変換する。**RM025** の全体反応は新しく導入されたカルボキシル基を側鎖部に残して効率的にアミノ酸の主鎖の一部を除去する。

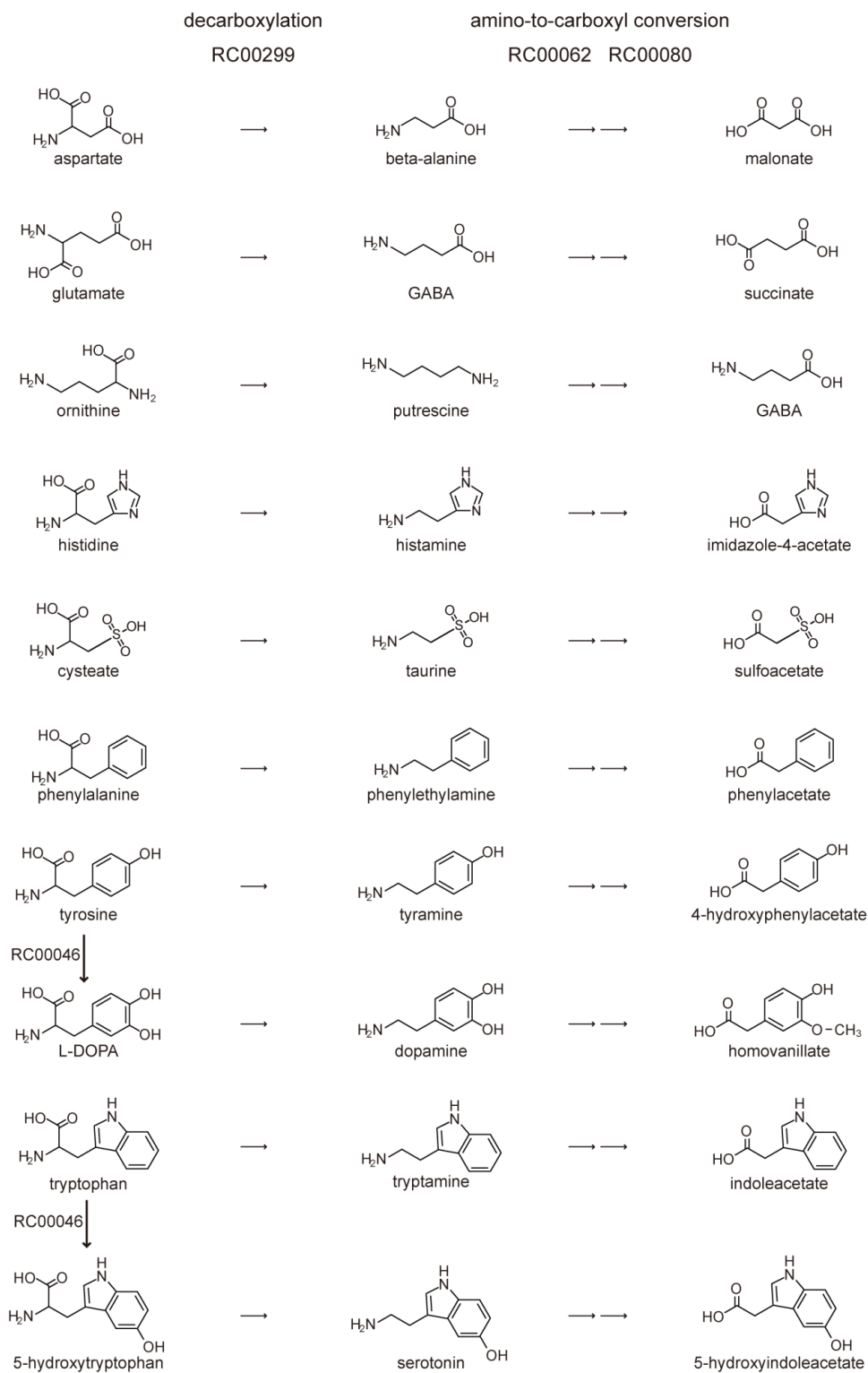


図 5-1 アミノ基-カルボキシル基変換モジュール

Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622](#). Copyright 2013 American Chemical Society.

代謝の進化における逆経路の利用については、これまで様々な研究で言及されている。1976 年の Jensen らによる総説では、逆方向の反応を触媒する酵素の遺伝子から遺伝子重複によって生まれたと考えられるホモログタンパク質の例が挙げられている。また、実験室において酵素活性を逆反応に転換させた研究についても言及している。1974 年の Cocks らの実験によって、嫌氣的条件下ではたらく L-lactaldehyde reductase が L-1,2-propanediol dehydrogenase に転換できることが示されている。

本研究で得られた逆経路の保存性は、生体内の化学変換における有機化学的制限が、逆反応においても成立する可能性を示唆している。また、代謝パスウェイの進化を考える際には、逆反応または逆経路を採用したことによる多様化についても考慮する必要がある。

5.2 モジュール間の包含関係

抽出された反応モジュールは、その多くが他のモジュールと包含関係にあるモジュールであった。代謝パスウェイ中の 5,805 の反応のうち約 60%が反応モジュールに対応することは既に述べたが、そのうち約 70%の反応は、他のモジュールを包含するモジュールに含まれていた。

4 章で述べたとおり、Caboxyl to amino conversion using protective N-acetyl group (RM002) と Caboxyl to amino conversion (RM032) の伸長モジュール(図 4-6)は興味深い知見を与えてくれる。RM002 の 5 反応のうち前後一反応ずつを除いた 3 反応は RM032 と一致する。つまり RM002 は RM032 を内含する反応モジュールであるが、アスパラギン酸のような短鎖分子に比べ、2-アミノアジピン酸のような長鎖分子では伸長反応は選択的に起こら

ず、分子内環化反応が自発的に起こってしまう。前の一反応はアセチル基をマスクすることでこの分子内環化を防ぎ、後の一反応は先に加えた N-アセチル基を除去することで分子修飾を外す。

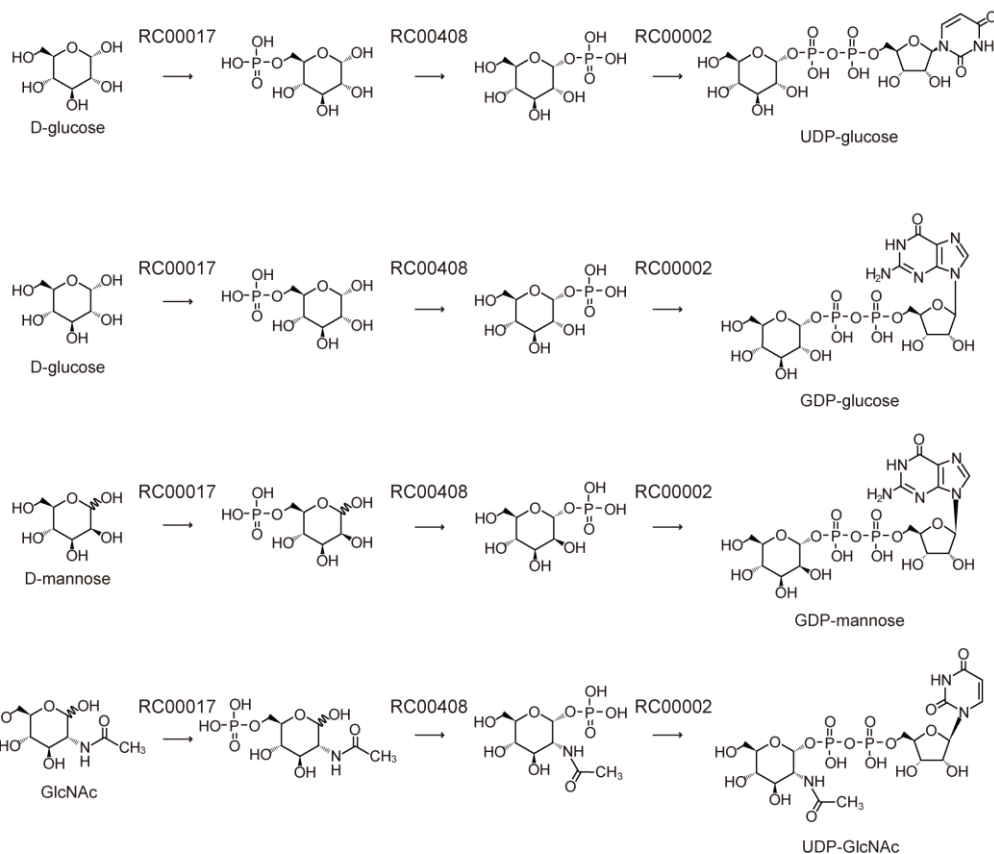
これらの例から、代謝パスウェイの形成においては、逆反応を含め様々なモジュールが組み合わされることで、有機反応の遷移状態の構造最適化が起こっていると考えられる。例えば、分子が大きく目的の反応が選択的に起こらない場合には、特異性を増大させるよう保護基を導入する。遷移状態までの活性化エネルギーが高すぎるようであれば、より低い活性化エネルギーで遷移状態を達成できる構造へ基質を変化させ、反応後に元の構造へ戻す。このような有機反応経路の最適化が代謝系の進化において淘汰圧として働いてきた結果、連続反応の保存性が生まれてきたものと考えられる。

5.3 リン酸化と配糖化

最も多くの反応ペアを含んでいた **RCLASS** エントリはリン酸やグリコシル化に関連したものであった。局所的な構造パターンに焦点を当てた我々のアプローチは変化する構造の複雑さに対する化学反応の進化を推察するのに成功したが、リン酸化やグリコシル化を含む反応モジュールを特徴づけるには、局所的な反応クラスに加え全体的な構造クラスを考慮する必要があるかもしれない。しかしながら、特定のクラスの全体構造がリン酸化配列のパターンから効率的に抽出できるケースも存在する。その一つの例は糖ヌクレオチドの生合成に見つけることができる。糖ヌクレオチドは多糖や糖化物質の生合成に用いられる単糖の活性型である。我々は二種類のリン酸反応モジュール **RM022** (**RC00017 RC00408 RC00002**)と **RM023** (**RC00078 RC00002**)を発見した。前者のモジュ-

ルは 6 位の炭素上にあったリン酸基が 1 位の炭素へと移動する異性化反応を含み、1 位の炭素上のリン酸基がヌクレオチドと結合される。(図 5-2) 後者のモジュールではリン酸基は直接 1 位に付き、RM022 と同様に 1 位の炭素上のリン酸基がヌクレオチドとの結合を受ける。

RM022



RM023

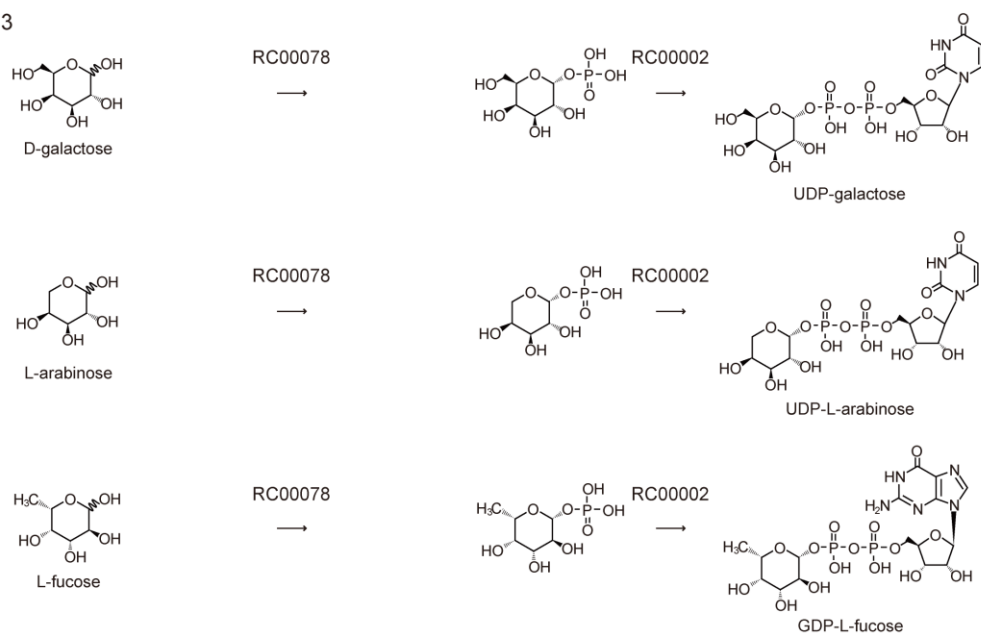


図 5-2 糖ヌクレオチド合成モジュール

Reprinted with permission from [J Chem Inf Model.](#), (2013)53, 613-622. Copyright 2013 American Chemical Society.

5.4 一次代謝と二次代謝のモジュール情報の相違

代謝ネットワークは、生命を維持するための保存された中心部分（原始代謝系）と環境と相互作用するための多様な表面部分（二次代謝系）からなると見ることもできる。微生物の生分解モジュールは様々な構造の生体異物を、一次代謝系の限られた数の化合物へと変換する。対照的に、特に植物二次代謝産物の生合成パスウェイは、限られた数の一次代謝化合物から多様な化合物を生成するパスウェイである。図 5-3 は、今回の手法で検出した短い反応モジュール RM027(表 4-2)を含む、KEGG(map00940)に示されるフェニルプロパノイド生合成パスウェイの中心部分を示したものである。反応モジュール RM027 はヒドロキシル化とメチル化反応からなり、シトクロム P450 モノオキシゲナーゼと *O*-メチルトランスフェラーゼに関連するよく知られた反応列である。このフェニルプロパノイド生合成パスウェイは、水平方向の RM027 の連続モジュールと、垂直方向の CoA が関わる酸-アルデヒド-アルコール変換のモジュールが交差するメッシュ様構造になっている。芳香環の 3,5-位の 2 か所でヒドロキシル化(RC00046)とメチル化(RC00392)の部分修飾が並列に起こる反応列であり、連続的な変化だけを検出する今回の手法では RM027 は検出できるが RM027 の 2 回繰り返しを 1 つのモジュールとして検出することが難しい。そのため、植物生合成パスウェイを解析した結果、生分解モジュールでは抽出できたような長いモジュールを発見することはできなかった。図 5-3 に示したような、同一分子上の異なる部分において反応が並列に起こる場合には、化合物を部分構造に分割して、部分構造毎の構造変換を検出することでパターン抽出ができる可能性がある [34]。植物二次代謝産物生合成パスウェイマップを詳細に検討すると、分子修飾に関与する短い化学反応モジュールが化合物上の複数の部位で並行に進行することで、多様な

化合物を生み出している例を見出すことができる。分子修飾モジュールに関与する可能性はある、P450 や *O*-メチルトランスフェラーゼに代表される修飾酵素は、植物では特に大きな遺伝子ファミリーを構成しており、植物はこれらの修飾酵素の多数のパラログを保有している。二次代謝生合成経路では、例に挙げたような、複数の反応部で並列に修飾が進行するモジュールの組み合わせが顕著に観察され、植物の多様な化合物合成能の一因となっていると考えられる。

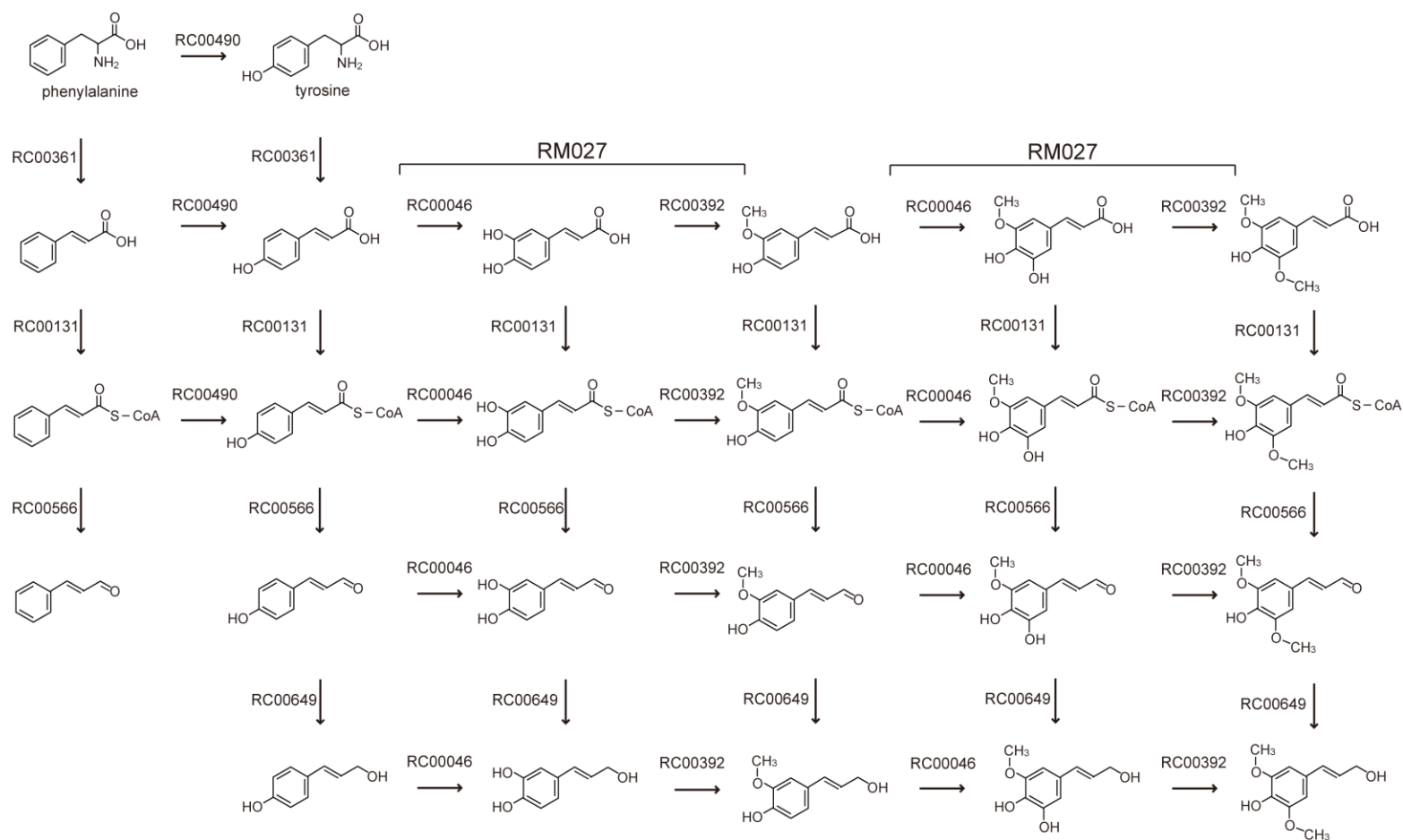


図 5-3 フェニルプロパノイド合成

水平方向の RM027 の連続モジュールと、垂直方向の CoA が関わる酸-アルデヒド-アルコール変換のモジュールが交差するメッシュ様構造。RM027 の連続モジュールは芳香族環上でヒドロキシル化(RC00046)とメチル化(RC00392)を行い、メトキシ基を生成する。CoA が関わる酸-アルデヒド-アルコール変換(RC00131+R00566+RC00649)はカルボン酸の還元を行う。Reprinted with permission from [J Chem Inf Model.](#) (2013)53, 613-622. Copyright 2013 American Chemical Society.

5.5 EC 番号による分類体系との比較

EC 体系は酵素反応の分類であり、4桁の番号のうち最初の3桁の数字は反応をクラス、sub クラス、sub-sub クラスへと階層的に分類する。4桁目の番号は基質特異性を区別するために付けられた番号である。KEGG には 8,990 の反応が蓄積されているが、本研究で用いた反応のうち EC 番号が与えられているものは 4,321 であった。EC 番号は生化学的に検証され文献に報告された反応にのみ与えられているため、EC 体系の対象となる範囲はより限られている。それに対し KEGG データベースは、反応が明らかになっているものに関しては、酵素や遺伝子が同定されていないものであっても収集・蓄積している。我々は EC 体系の中でもっとも細かい反応分類である EC sub-sub クラスを KEGG RCLASS(RC)と比較した。EC sub-sub クラス 249 に対し 2,481 の RC エントリの対応を調べたところ、平均して 13 の RC エントリが KEGG パスウェイ上の EC sub-sub クラスに対応していた。明らかに、RC 体系は EC 体系よりも細分化された反応分類となっている。表 5-1 中に最も多い RC エントリに対応した EC sub-sub クラスの一覧を示す。

1 章で述べたとおり、一部の EC 階層においては、主産物の有機反応の分類よりも補酵素の一致などが優先され、必ずしも有機反応の類似性が反映されていない (図 1-5)。例えば、CYP モノオキシゲナーゼの反応は有機反応としてはヒドロキシラーゼ、エポキシダーゼ (エポキシゲナーゼ)、ベイヤー・ビリガー モノオキシゲナーゼその他に分類することができるが、これらの有機反応分類と EC 体系は一致していない。本研究で定義した反応の類似性指標を使い、これらの反応クラスをグルーピングしたところ、類似性スコアの閾値=0.9の時に8~40の反応クラスを2~6のグループにまとめることが出来た。

(図 5-4i)。また、1 章で挙げた、基質と産物の有機反応は類似しているにも拘らず、補

酵素の違いによって異なる EC sub-sub classes に分類されたヒドロキシラーゼ反応(図 1-5a)、反応中心となる原子種が異なるにも拘らず、補酵素が同一であるために同一の EC sub-sub classes に分類されたヒドロキシラーゼ反応(図 1-5a)、についても、RCLASS では有機反応により分類できていることを図 5-4(ii), (iii)に示す。

表 5-1 最大の RC エントリを包含する EC 番号

Reprinted with permission from [J Chem Inf Model., \(2013\)53, 613-622.](#) Copyright 2013 American Chemical Society.

EC sub-sub class	# of RC entries	Enzymes involved
1.14.13	150	monooxygenases
4.2.1	129	hydratase/dehydratases, terpenecyclases (hydrating)
1.1.1	117	alcohol,NAD(P)+ dehydrogenases
4.2.3	86	phospho-lyases, terpenecyclases (diphosphate-eliminating)
1.14.14	83	monooxygenases
2.1.1	79	methyltransferases
1.13.11	72	dioxygenases
1.3.1	68	saturases/desaturases
4.1.1	65	carboxylases/decarboxylases
2.5.1	60	prenyltransferases, 1-carboxyvinyltransferases,aminocarboxyethyltransferases, aminocarboxypropyltransferases,adenosyltransferases

(i) Number of RClass entries (or similarity groups of RClass entries)
corresponding to various types of monooxygenases

reaction	ec1.14.13			ec1.14.14			ec1.14.15		
	RC	RG1.0	RG0.9	RC	RG1.0	RG0.9	RC	RG1.0	RG0.9
hydroxylase	40	19	6	21	14	4	8	4	2
epoxidase	8	4	2	13	9	2			
BVMO	14	6	4						
other	77	63	48	46	40	30	10	8	5

*RG: Group in which RClass entries were grouped by single linkage

(ii) Some type of hydroxylase reactions in different EC sub-sub classes

EC: 1.14.13	<p>R05736 (EC: 1.14.13.72)</p>	RC00257
EC: 1.14.14	<p>R01840 (EC: 1.14.14.1)</p>	
EC: 1.14.15	<p>R03262 (EC: 1.14.15.5)</p>	

(iii) Various types of hydroxylase reactions found in an EC sub-subclass

EC: 1.14.13	<p>R00815 (EC: 1.14.13.7)</p>	RC00236
	<p>R09861 (EC: 1.14.13.143)</p>	RC00478
	<p>R00448 (EC: 1.14.13.59)</p>	RC00298

図 5-4 モノオキシゲナーゼの EC 分類と RCLASS の比較

(i) CYP モノオキシゲナーゼの反応クラスの EC sub-sub クラス分類と RCLASS 類似性スコアによるグルーピングの結果。(ii) 基質と産物の有機反応は類似しているにも拘らず、補酵素の違いによって異なる EC sub-sub classes に分類されたヒドロキシラーゼ反応と対応する RCLASS。(iii) 反応中心となる原子種が異なるにも拘らず、補酵素が同一であるために同一の EC sub-sub classes に分類されたヒドロキシラーゼ反応と、対応する RCLASS。Reprinted with permission from [J Chem Inf Model.](#), (2013)53, 613-622. Copyright 2013 American Chemical Society.

5.6 モジュール情報の応用

遺伝的ネットワークと化学的ネットワークという代謝系についての二つの見方には実用上の価値がある。これまで遺伝子やタンパク質に関する知識から化合物や反応の予測が行われてきたが、化合物や反応についての知識の蓄積によって、これまでとは逆に反応から遺伝子やタンパク質を予測することも可能であると考ええる。KEGG データベースでは、全ての利用可能な完全ゲノムについて既知遺伝子のオーソログを探索した結果が KEGG オーソロジー(KO)システムによってカテゴリ分けされており、生命のゲノム空間の理解を目指している。本研究では、生命とそれを取り巻く環境の化学空間を理解する試みとして、反応クラス (RC) 体系によって、全ての既知の反応のカテゴリ分けを行った。我々は、特に代謝ネットワークのモジュール構造について、KO 体系と RC 体系の関連性の検出を試みている。化学反応モジュールとして抽出された代謝ネットワーク上の生化学的禁制律と、生物の持つオーソログ情報の両方を考慮することにより、より正確な代謝パスウェイ予測が可能になると考える。本研究で得られた化学反応モジュールは、KEGG データベース (<http://www.genome.jp/kegg/>) の一部として Reaction Module (<http://www.genome.jp/kegg/reaction/rmodule.html>) に公開されている。

5.7 代謝ネットワークの進化

これまで提唱されてきた代謝ネットワークの進化モデルは、酵素の進化として代謝進化を表してきた。しかしながら、酵素の進化と反応の進化は代謝系の進化の両輪であり、代謝ネットワークの進化を考える際には、酵素のネットワークだけではなく化学的ネッ

ネットワークの進化という観点に基づいても考える必要がある。本研究で抽出した化学反応モジュールは、酵素のアミノ酸配列非依存に、反応パターンのみに基づいた類似性指標によって定義されている。よって代謝ネットワークを化学的ネットワークとして捉え、反応列の多様性によって網羅的に代謝ネットワークからの特徴抽出を行った初めての例である。

抽出された反応モジュールは、その多くが他のモジュールと包含関係にあるモジュールであった。代謝パスウェイ中の反応 5,805 のうち、およそ 60%が反応モジュールに含まれ、そのうち約 70%にあたる反応が、他のモジュールを包含するモジュールに含まれていた。また、短いモジュールは幅広い代謝系から抽出され、抽出される数も多いのに対し、モジュールが長いほど抽出される数は少なく、かつ限られた代謝系のみから抽出された(図 4-1)。このことから、短いモジュールほど全代謝系の中で一般的な反応列であり、長いモジュールになるほど一部の代謝系に特異的に出現する反応列であることがわかる。

4 章で述べたとおり、Caboxyl to amino conversion using protective N-acetyl group (RM002) と Caboxyl to amino conversion (RM032) の伸長モジュール(図 4-6)は興味深い知見を与えてくれる。RM002 の 5 反応のうち前後一反応ずつを除いた 3 反応は RM032 と一致する。つまり RM002 は RM032 を包含する反応モジュールである。アスパラギン酸のような短鎖分子に比べ、2-アミノアジピン酸のような長鎖分子では伸長反応は選択的に起こらず、自発的に分子内環化反応が起こってしまうことで伸長反応が妨げられる。RM002 では前の一反応が N-アセチル基を付加し環化反応を起こす官能基をマスクすることで分子内環化を防ぎ、後の一反応が先に加えた N-アセチル基を除去することで分子修飾を外す。

よって、RM002 はコアとなる RM032 の反応列に、それらを補助する反応が加わったモジュールであると言える。

このようなコア反応を補助する反応が加わった反応モジュールの存在は、代謝ネットワークの進化に関して次のような考察を与える。反応ネットワークの進化として代謝の進化を考えると、化学反応モジュールが抽出された経路は共通の反応列からなる姉妹パスウェイである。RM002 のような、コア反応を補助する反応が加わった反応モジュールが検出されることは、それまでコア反応の対象とならなかった基質に対し、酵素の進化だけでなく代謝パスウェイが新たな反応列を形成することによってコア反応を行うことが可能になった可能性を示唆する。またこのモジュールが複数の代謝パスウェイから検出されることから、それ自体が単位となって新たな姉妹パスウェイが生じていることを示す。これらのことから、反応列の多様化に基づく代謝ネットワークの進化には、短い反応列のコアに対しそれを補助する反応が付け加わっていくことにより、化学反応ネットワークが拡大してきたというモデルが考えられる。これを酵素の進化に置き換えると、まず、酵素の基質非特異性によって、ある基質を変換していた酵素群が新規の基質にも同じ反応を行うようになる。但しこのとき生じた新規経路は特異性の低い非効率な反応列である。これらの非効率的な反応列が効率的な経路へと変化するには以下の 2 つの可能性が考えられる。一つ目は酵素の進化である。遺伝子重複などが起こることにより、以前の基質への特異性に縛られず新規基質のみに適応すれば、それらの酵素群は高い効率で新規基質を変換できるようになる。もう一つは反応列の進化である。別の経路で使われていた酵素が、新規経路が反応を特異的に成しうるよう補助的な役割を果たすこと

で、コア反応が効率よく起こるようになる。こうして出来上がった新規経路は、新規基質に特化した結果、旧経路では対象になり得なかった新たな基質を対象にすることができるようになるかもしれない。この繰り返しによって、代謝パスウェイはより長く、より多様になってきたという進化の仮説が考えられる。このモデルでは従来の **patchwork model**[10]を踏襲しつつ、化学ネットワークの進化についての観点を取り入れている。

第6章 総括

本研究では、代謝パスウェイデータベースから保存された連続反応のパターンを抽出し、代謝の機能単位と考えられるものに機能アノテーションを与え、化学反応モジュールとしてデータベース化した。これらの化学反応モジュールから、実用的にも理学的にも興味深い特徴が得られた。第一に、化学反応モジュールは異なる化学物質を生成する異なるパスウェイで繰り返し出現する。第二に、化学反応モジュールは代謝ネットワークの基本構成要素であるかのように組み合わせられて用いられている。第三に、反応モジュールはゲノム上の酵素のオーソログ遺伝子クラスターにより定義された KEGG パスウェイモジュール、特に酵素群をコードするオペロン様構造の遺伝子クラスターと対応する傾向がある。

また、抽出された反応モジュールは、その多くが他のモジュールと包含関係にあるモジュールであった。代謝パスウェイ中の反応 5,805 のうち、およそ 60%が反応モジュールに対応し、そのうち約 70%にあたる反応が、他のモジュールを包含するモジュールに含まれていた。また、短いモジュールは幅広い代謝系から抽出され、抽出される数も多いのに対し、モジュールの長さが長いほど抽出される数は少なく、かつ限られた代謝系のみから抽出された。このことから、短いモジュールほど全代謝系の中で一般的な反応列であり、長いモジュールになるほど一部の代謝系に特異的に出現する反応列であることがわかる。本研究ではこの傾向から、コア反応に補助的な反応が加わることの繰り返しによって、代謝パスウェイは伸長し多様化してきたという進化の仮説を考えた。この

仮説はこれまでの酵素の進化という観点に加え、有機化学的禁制律に基づいた代謝反応の進化という観点を提案するものである。

本研究において得られた化学反応モジュール情報は、genome-driven な代謝パスウェイ再構築手法による生体機能予測や、新規化合物に対する合成・分解経路の予測に有力な材料となると考えられる。本研究で得られた反応モジュールは、KEGG データベース (<http://www.genome.jp/kegg/>)の一部として公開されている。

謝辞

本研究は、京都大学化学研究所バイオインフォマティクスセンター生命知識システム領域（現 化学生命科学研究領域）において、金久實教授の指導のもとに行われました。本研究を遂行するにあたり、長期に渡りご指導頂きました金久教授に心より感謝致します。同研究室の五斗進准教授には結果の考察や論文の推敲において細やかなご指導を頂きました。小寺正明助教には日頃から多くの助言と激励を頂き、論文の推敲に関しご助力を頂きました。時松敏明助教には結果の解釈に関し助言を頂き、特に二次代謝経路に関して貴重なご意見を頂きました。中川善一研究員にはデータ整備を行なって頂き、結果の解釈に関し貴重なご意見を下さいました。皆様に深く感謝の意を表したく思います。また、日頃から様々な面でお世話になりました秘書の皆様、活発な議論によって多くの発想を与えてくださった研究室の皆様にも感謝申し上げます。最後に、長きにわたり研究生生活を支えて下さった両親、家族、友人の皆様にも心より感謝致します。

参考文献

- [1] 榊原圭子、斉藤和季：メタボロミクスを基盤とした植物ゲノム機能科学、メタボロミクスの先端技術と応用 pp.204-211、CMC 出版、2008
- [2] McDonald, A.G.; Boyce, S.; Tipton, K.F. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res.* **2009**, 37, D593–D597.
- [3] Orengo, C.A. et al.; (1997) CATH – a hierarchic classification of protein domain structures. *Structure* 5, 1093–1108
- [4] Murzin, A.G. et al. (1995) SCOP: a structural classification of proteins database for the
- [5] Schmidt S, Sunyaev S, Bork P, Dandekar T.; Metabolites: a helping hand for pathway evolution? *Trends Biochem Sci.* 2003 Jun;28(6):336-41.
- [6] Horowitz NH.; On the Evolution of Biochemical Syntheses. *Proc Natl Acad Sci U S A.* 1945 Jun;31(6):153-7.
- [7] Horowitz NH., Netzenberg RL.; Biochemical Aspects of Genetics. *Annu Rev Biochem.* 1965;34:527-64.
- [8] Lawrence JG, Roth JR.; Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics.* 1996 Aug;143(4):1843-60.
- [9] Siefert JL, Martin KA, Abdi F, Widger WR, Fox GE.; Conserved gene clusters in bacterial genomes provide further support for the primacy of RNA. *J Mol Evol.* 1997 Nov;45(5):467-72.

- [10]Jensen RA.; Enzyme recruitment in evolution of new function. *Annu Rev Microbiol.* 1976;30:409-25.
- [11]Anbar AD.; Elements and evolution. *Science.* 2008 Dec 5;322(5907):1481-3. doi: 10.1126/science.1163100.
- [12]Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2012**, 40, D109-D114.
- [13] Bono, H.; Ogata, H.; Goto, S.; Kanehisa, M. Reconstruction of amino acid biosynthesis pathways from the complete genome sequence. *Genome Res.* **1998**, 8, 203-210.
- [14]Galperin, M.Y.; Koonin, E.V. Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. *Genetica* **1999**, 106, 159-170.
- [15]Dandekar, T.; Schuster, S.; Snel, B.; Huynen, M.; Bork, P. Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem. J.* **1999**, 343, 115-124.
- [16]Forst, C.V.; Schulten, K. Evolution of metabolisms: a new method for the comparison of metabolic pathways using genomics information. *J.Comput. Biol.* **1999**, 6, 343-360.
- [17]Ravasz, E.; Somera, A.L.; Mongru, D.A.; Oltvai, Z.N.; Barabási, A.L. Hierarchical organization of modularity in metabolic networks. *Science* **2002**, 297, 1551-1555.
- [18]Papin JA, Reed JL, Palsson BO.; Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends Biochem Sci.* 2004 Dec;29(12):641-7.
- [19]Tohsato, Y.; Matsuda, H.; Hashimoto, A. A multiple alignment algorithm for metabolic pathway analysis using enzyme hierarchy. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2000**, 8, 376-383.

- [20] Pinter, R.Y.; Rokhlenko, O.; Yeager-Lotem, E.; Ziv-Ukelson, M. Alignment of metabolic pathways. *Bioinformatics***2005**, 21, 3401-3408.
- [21] Wernicke, S.; Rasche, F. Simple and fast alignment of metabolic pathways by exploiting local diversity. *Bioinformatics***2007**, 23, 1978-1985.
- [22] Ay, Y.; Kellis, M.; Kahveci, T. SubMAP: aligning metabolic pathways with subnetwork mappings. *J. Comput. Biol.***2011**, 18, 219–235.
- [23] Hattori, M.; Okuno, Y.; Goto, S.; Kanehisa, M. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.***2003**, 125, 11853-11865.
- [24] Kotera, M.; Okuno, Y.; Hattori, M.; Goto, S.; Kanehisa, M. Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *J. Am. Chem. Soc.***2004**, 126, 16487-16498.
- [25] Horie, A.; Tomita, T.; Saiki, A.; Kono, H.; Taka, H.; Mineki, R.; Fujimura, T.; Nishiyama, C.; Kuzuyama, T.; Nishiyama, M. Discovery of proteinaceous N-modification in lysine biosynthesis of *Thermoterrivibrio*. *Nat. Chem. Biol.***2009**, 5, 673-679.
- [26] Pelletier, D.A.; Harwood, C.S. 2-Hydroxycyclohexanecarboxyl coenzyme A dehydrogenase, an enzyme characteristic of the anaerobic benzoate degradation pathway used by *Rhodospseudomonas palustris*. *J. Bacteriol.***2000**, 182, 2753-2760.
- [27] Harwood, C.S.; Parales, R.E. The beta-ketoadipate pathway and the biology of self-identity. *Annu. Rev. Microbiol.* **1996**, 50, 553-590.

- [28] Rabus, R.; Kube, M.; Heider, J.; Beck, A.; Heitmann, K.; Widdel, F.; Reinhardt, R. The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. *Arch. Microbiol.* **2005**, 183, 27-36.
- [29] Lee, S.H.; Jin, H.M.; Lee, H.J.; Kim, J.M.; Jeon, C.O. Complete genome sequence of the BTEX-degrading bacterium *Pseudoxanthomonas padix* BD-a59. *J. Bacteriol.* **2012**, 194, 544.
- [30] Maeder, D.L.; Weiss, R.B.; Dunn, D.M.; Cherry, J.L.; González, J.M.; DiRuggiero, J.; Robb, F.T. Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* **1999**, 152, 1299-1305.
- [31] Papin, J.A.; Reed, J.L.; Palsson, B.O. Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends Biochem. Sci.* **2004**, 29, 641-647.
- [32] Schuster, S.; Pfeiffer, T.; Moldenhauer, F.; Koch, I.; Dandekar, T. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics* **2002**, 18, 351-361.
- [33] Yamada, T.; Kanehisa, M.; Goto, S. Extraction of phylogenetic network modules from the metabolic network. *BMC Bioinformatics* **2006**, 7, 130.
- [34] Chu, H.Y.; Wegel, E.; Osbourn, A. From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. *Plant J.* **2011**, 66, 66-79.